

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 19:17:09 ; Search time 48056 Seconds
(without alignments)
7167.048 Million cell updates/sec

Title: US-10-673-885-1
Perfect score: 7108
Sequence: 1 taattgtacttgcagaa.....gcttctcactgtgcttt 7108

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_btg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7108	100.0	7108	6 AR527423	AR527423 Sequence
2	6800.8	95.7	6897	6 AX751409	AX751409 Sequence
3	6799.8	95.7	7225	6 AX644236	AX644236 Sequence
4	6683.6	94.0	6903	6 AX686453	AX686453 Sequence
5	6453	90.8	6994	6 AX686449	AX686449 Sequence
6	5619.4	79.1	5877	6 AX751415	AX751415 Sequence
7	5294.2	74.5	7396	10 AF063249	AF063249 Rattus no
8	2618.6	36.8	3973	6 AR073855	AR073855 Sequence
9	2618.6	36.8	3973	6 I32039	I32039 Sequence 21
10	2551	35.9	2565	6 AX686451	AX686451 Sequence
11	2251	31.7	3149	6 CQ726350	CQ726350 Sequence
12	1118.8	15.7	2692	6 AR073854	AR073854 Sequence
13	1118.8	15.7	2692	6 I32038	I32038 Sequence 14
14	1071.6	15.1	2309	6 AR073853	AR073853 Sequence
15	1071.6	15.1	2309	6 I32037	I32037 Sequence 10
16	905.2	12.7	990	6 AX751414	AX751414 Sequence
17	789.6	11.1	2173	6 AR073851	AR073851 Sequence
18	789.6	11.1	2173	6 I32035	I32035 Sequence 6
19	675.8	9.5	834	6 CQ731368	CQ731368 Sequence

20	532	7.5	532	6 AR073852	AR073852 Sequence
21	532	7.5	532	6 I32036	I32036 Sequence 8
22	401.6	5.6	402	9 AF169351	AF169351 Homo sapi
23	330.4	4.6	163912	9 AC074031	AC074031 Homo sapi
24	329.2	4.6	659	6 CQ723010	CQ723010 Sequence
25	319.4	4.5	321	6 AR073850	AR073850 Sequence
26	319.4	4.5	321	6 I32034	I32034 Sequence 3
27	315	4.4	254366	6 AR527424	AR527424 Sequence
28	285.6	4.0	179162	2 AC132076	AC132076 Bos tauru
29	285.6	4.0	236486	2 AC138166	AC138166 Bos tauru
30	283	4.0	162738	9 AC083812	AC083812 Homo sapi
31	277	3.9	174126	2 AC135552	AC135552 Canis fam
32	276	3.9	141299	2 AC132080	AC132080 Felis cat
33	267.4	3.8	154909	2 AC136206	AC136206 Sus scrof
34	266.8	3.8	363	6 CQ736545	CQ736545 Sequence
35	251.2	3.5	162738	9 AC083812	AC083812 Homo sapi
36	247.2	3.5	230799	2 AC098397	AC098397 Rattus no
37	245.6	3.5	203839	10 AC021642	AC021642 Mus muscu
38	230.8	3.2	412	6 AX751416	AX751416 Sequence
39	225.6	3.2	164493	2 AC078825	AC078825 Homo sapi
40	214.6	3.0	4053	5 GGU65891	UG5891 Gallus gall
41	214	3.0	214819	10 AC123948	AC123948 Mus muscu
42	201.6	2.8	261708	2 AC098350	AC098350 Rattus no
43	195.2	2.7	3581	4 OC032587	U32587 Oryctolagus
44	195.2	2.7	5704	4 OC009490	U09490 Oryctolagus
45	194.6	2.7	5515	5 GGA238216	AJ238216 Gallus ga

ALIGNMENTS

RESULT 1
AR527423
LOCUS AR527423 7108 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 1 from patent US 6723547.
ACCESSION AR527423
VERSION AR527423.1 GI:53914473
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7108)
Beasley,E.M., Webster,M., Francesco,V.D. and Wei,M.-H.
TITLE Isolated human phosphatase proteins, nucleic acid molecules
encoding human phosphatase proteins, and uses thereof
JOURNAL Patent: US 6723547-A 1 20-APR-2004;
FEATURES Location/Qualifiers
source
1..7108
/organism="unknown"
/mol_type="genomic DNA"

Query Match	Best Local Similarity	Matches 7108;	Conservative	Score 7108;	DB 6;	Length 7108;	Mismatches	Indels	Gaps
100.0%	100.0%	0	0	0	0	0	0	0	0
QY	1	TAATTGTGACTTGCAGAGGATCTGCTTTAAATCATTAATGACGAGCAACATTCTCT	60						
DB	1	TAATTGTGACTTGCAGAGGATCTGCTTTAAATCATTAATGACGAGCAACATTCTCT	60						
QY	61	CTAGAGCCATCAATGTGATTTCTACTGCGTGAATAATGTAATAAGATGGATTTCTTATC	120						
DB	61	CTAGAGCCATCAATGTGATTTCTACTGCGTGAATAATGTAATAAGATGGATTTCTTATC	120						
QY	121	ATTTTCTTTTACTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCATGTCGTT	180						
DB	121	ATTTTCTTTTACTTTTATTTGGACTTCAGAGACACAGGTTGATGTTTCCATGTCGTT	180						
QY	181	CCTGGTACTAGGTACGATATAACCATCTCTCAATTTCTACACATACACCTCACCTGTT	240						
DB	181	CCTGGTACTAGGTACGATATAACCATCTCTCAATTTCTACACATACACCTCACCTGTT	240						
QY	241	ACTAGAATAGTGACACCAAAATGTAAACAAACGAGGCGCTCCAGTCTTCTTAGCCGGGAA	300						

Qy	2461	CCCAATGGAAATCATAAAAAATAATATACAATTTATCTCAAGAGAAAGTAATGGAATATGAGAA	2520
Db	2461	CCCAATGGAAATCATAAAAAATAATATACAATTTATCTCAAGAGAAAGTAATGGAATATGAGAA	2520
Qy	2521	AGAACTATAAATACAAACCTCTTTAAACCCAAAAACATTAAGGTACTGGAAGAATAATACC	2580
Db	2521	AGAACTATAAATACAAACCTCTTTAAACCCAAAAACATTAAGGTACTGGAAGAATAATACC	2580
Qy	2581	TATATCAATGAGGTGTCTGTAGTACATCAAGAAAGTGAAGAGTTCCGAGTGTCTCCCA	2640
Db	2581	TATATCAATGAGGTGTCTGTAGTACATCAAGAAAGTGAAGAGTTCCGAGTGTCTCCCA	2640
Qy	2641	AGTATACCTGACGGAGGAGATGCTCTGATTTCTCCCCCTCAAGACTTCTCTGTAAACAG	2700
Db	2641	AGTATACCTGACGGAGGAGATGCTCTGATTTCTCCCCCTCAAGACTTCTCTGTAAACAG	2700
Qy	2701	TTGTCTGGTGTACGGTGAAGTTGTCTATGGCAACACCCCTCGAGGCCAAATGGAATTTATC	2760
Db	2701	TTGTCTGGTGTACGGTGAAGTTGTCTATGGCAACACCCCTCGAGGCCAAATGGAATTTATC	2760
Qy	2761	CTTTATTAACACAGTTTATGTCTGGAATAGATCATCTATTAATAAACTATTAATGTCACTGAA	2820
Db	2761	CTTTATTAACACAGTTTATGTCTGGAATAGATCATCTATTAATAAACTATTAATGTCACTGAA	2820
Qy	2821	ACATCATTTGGAGTTATCAGATTTGGATTATAATGTTGAAATACAGTGTCTATGTAACAGCT	2880
Db	2821	ACATCATTTGGAGTTATCAGATTTGGATTATAATGTTGAAATACAGTGTCTATGTAACAGCT	2880
Qy	2881	AGCACACAGATTTGGTGTATGGGAAACAGGAAGCAATATCAATTAGCTTTCAAAACACAGAG	2940
Db	2881	AGCACACAGATTTGGTGTATGGGAAACAGGAAGCAATATCAATTAGCTTTCAAAACACAGAG	2940
Qy	2941	GGAGACCAAGCGATCTCCCAAAGATGTTTATATGCAAAACCTCAGTCTTCAATCAATA	3000
Db	2941	GGAGACCAAGCGATCTCCCAAAGATGTTTATATGCAAAACCTCAGTCTTCAATCAATA	3000
Qy	3001	ATTCTTTTCTGGACACCTCTCCAAACCTAAATGGGATTAACAATATTACTCTGTTTAT	3060
Db	3001	ATTCTTTTCTGGACACCTCTCCAAACCTAAATGGGATTAACAATATTACTCTGTTTAT	3060
Qy	3061	TACAGAAATACTTCAGGTACTTTTATGCAAGAAATTTTACATCCATGAACTAAACCAATGAC	3120
Db	3061	TACAGAAATACTTCAGGTACTTTTATGCAAGAAATTTTACATCCATGAACTAAACCAATGAC	3120
Qy	3121	TTTGCAAAATAGCTGTATCCAAATPATAGATAAATCTGACAAATPATTCAGTACTATPACA	3180
Db	3121	TTTGCAAAATAGCTGTATCCAAATPATAGATAAATCTGACAAATPATTCAGTACTATPACA	3180
Qy	3181	TTTTGGTTTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAAATGAAGAGTGTGATCATTTGAA	3240
Db	3181	TTTTGGTTTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAAATGAAGAGTGTGATCATTTGAA	3240
Qy	3241	GTATPACACAGATCAAGACATACCTGGAAGGGTTGTTTGGAAACCTGACTTACGAATCCATT	3300
Db	3241	GTATPACACAGATCAAGACATACCTGGAAGGGTTGTTTGGAAACCTGACTTACGAATCCATT	3300
Qy	3301	TCGTCAAATGGAATAAATGTAAAGCTGGGTCCACCGGCTCAACCAACGGTGTAGTCTTC	3360
Db	3301	TCGTCAAATGGAATAAATGTAAAGCTGGGTCCACCGGCTCAACCAACGGTGTAGTCTTC	3360
Qy	3361	TACTATGTTTCACTGATCTTACAGCAGACTCTCTCGCCATGTGAGACCACCTCTTGTTTACA	3420
Db	3361	TACTATGTTTCACTGATCTTACAGCAGACTCTCTCGCCATGTGAGACCACCTCTTGTTTACA	3420
Qy	3421	TATGAGAGAAAGCATATATTTTGTAAATCTCGAAAAATAACACTGATTATATATTTAAAAAT	3480
Db	3421	TATGAGAGAAAGCATATATTTTGTAAATCTCGAAAAATAACACTGATTATATATTTAAAAAT	3480
Qy	3481	ACTCCATCAACAGAAAAAGGATTTCTCTGATACCTATATCTGCCAGCTATACATCAAGACT	3540
Db	3481	ACTCCATCAACAGAAAAAGGATTTCTCTGATACCTATATCTGCCAGCTATACATCAAGACT	3540

QY	3541	GAAGAAGATGTC	CCAGAAACCTT	CACCAATAT	CAACACTTTT	AAAAACCTT	CTCTACC	3600
DB	3541	GAAGAAGATGT	CCAGAAACCTT	CACCAATAT	CAACACTTTT	AAAAACCTT	CTCTACC	3600
QY	3601	TCAGTCTCTT	TATCATCGGAT	CCCCAGTAA	GCCAAATG	CGTCAATAA	TAAGTTATGAT	3660
DB	3601	TCAGTCTCTT	TATCATCGGAT	CCCCAGTAA	GCCAAATG	CGTCAATAA	TAAGTTATGAT	3660
QY	3661	TTAACTTTT	CAAGGACCAAA	TGAAAATTA	TTCTTTCATTA	CTTCTGATAA	TACATAATA	3720
DB	3661	TTAACTTTT	CAAGGACCAAA	TGAAAATTA	TTCTTTCATTA	CTTCTGATAA	TACATAATA	3720
QY	3721	TTGGAAGAGCT	TTTCCACCAATTA	CATTATAGCT	TTTTTGTCTG	CCGCAAGACTA	GAGAAAA	3780
DB	3721	TTGGAAGAGCT	TTTCCACCAATTA	CATTATAGCT	TTTTTGTCTG	CCGCAAGACTA	GAGAAAA	3780
QY	3781	GGACTTGT	CTCTCCAGAT	TTCTTTTCTTTT	TACACAGAT	GAGTCAGT	GCCTGTAGCACCT	3840
DB	3781	GGACTTGT	CTCTCCAGAT	TTCTTTTCTTTT	TACACAGAT	GAGTCAGT	GCCTGTAGCACCT	3840
QY	3841	CCACAAAAT	TTGACTTTAA	CTGACTTT	CAGACTTT	TGATGGCT	GAAATGGAGCCCA	3900
DB	3841	CCACAAAAT	TTGACTTTAA	CTGACTTT	CAGACTTT	TGATGGCT	GAAATGGAGCCCA	3900
QY	3901	AGTCCTCT	CCAGGTGTA	TGTTAAAGT	ATATAGTTT	AAAAATTCAT	GAAATGGAACCT	3960
DB	3901	AGTCCTCT	CCAGGTGTA	TGTTAAAGT	ATATAGTTT	AAAAATTCAT	GAAATGGAACCT	3960
QY	3961	GACACTAT	ATATTAAGA	ATATATCAGGA	TTTAAAC	TGAAGCCAA	ACTTGTGTGACTG	4020
DB	3961	GACACTAT	ATATTAAGA	ATATATCAGGA	TTTAAAC	TGAAGCCAA	ACTTGTGTGACTG	4020
QY	4021	GAAACAGT	CAGCACCT	ACTCTAT	CTCGTGTAT	CTCGCTTCA	CCAAAGTTGGAAATGGCAAT	4080
DB	4021	GAAACAGT	CAGCACCT	ACTCTAT	CTCGTGTAT	CTCGCTTCA	CCAAAGTTGGAAATGGCAAT	4080
QY	4081	CAATTTAGT	TAATGTAGT	AAAAATTC	CAACCCAA	AGAATCAGT	TTCCAGATGTCGTGCAAGAT	4140
DB	4081	CAATTTAGT	TAATGTAGT	AAAAATTC	CAACCCAA	AGAATCAGT	TTCCAGATGTCGTGCAAGAT	4140
QY	4141	ATGCAGT	GCATGCG	CACTAGCT	GGCAGT	CAGTTTATG	TGAAATGGGATCCACCCCAAAAG	4200
DB	4141	ATGCAGT	GCATGCG	CACTAGCT	GGCAGT	CAGTTTATG	TGAAATGGGATCCACCCCAAAAG	4200
QY	4201	GCAAAATG	GAATAATA	TAACG	CAGTATATG	GTAAACAGT	TTCTACAAAAGTTTCT	4260
DB	4201	GCAAAATG	GAATAATA	TAACG	CAGTATATG	GTAAACAGT	TTCTACAAAAGTTTCT	4260
QY	4261	CCCCAAGAT	CACATGTA	CACTTTTCA	TAAAGCTTCT	TGGCCAA	TACCTCATATGCTTTTAA	4320
DB	4261	CCCCAAGAT	CACATGTA	CACTTTTCA	TAAAGCTTCT	TGGCCAA	TACCTCATATGCTTTTAA	4320
QY	4321	GTAAGAGCT	TTCAACCT	CAGCTG	GTGAAGTGAT	GAAAGCACA	TGTCAGCACATA	4380
DB	4321	GTAAGAGCT	TTCAACCT	CAGCTG	GTGAAGTGAT	GAAAGCACA	TGTCAGCACATA	4380
QY	4381	CCTGAAAAC	AGTCTCC	AGTGTTC	CCCAAAATAT	TGCTTTTCT	CAATGTCAGTCAACTAGT	4440
DB	4381	CCTGAAAAC	AGTCTCC	AGTGTTC	CCCAAAATAT	TGCTTTTCT	CAATGTCAGTCAACTAGT	4440
QY	4441	GCAACATT	GATGGA	TGAAGCT	GCACATAT	CTTTGGCT	ACTTTTCAAAATTA	4500
DB	4441	GCAACATT	GATGGA	TGAAGCT	GCACATAT	CTTTGGCT	ACTTTTCAAAATTA	4500
QY	4501	ACGACTCA	ACTTCGT	GTCTAAAA	ATGGAAGAA	TGGGAATCC	GAAGATGTGTGTAATAT	4560
DB	4501	ACGACTCA	ACTTCGT	GTCTAAAA	ATGGAAGAA	TGGGAATCC	GAAGATGTGTGTAATAT	4560
QY	4561	CAAAAAAT	TCAAAT	CTCTAT	GAGCTCACTT	AACTGA	GAGACAGTATATGGAATTAAG	4620
DB	4561	CAAAAAAT	TCAAAT	CTCTAT	GAGCTCACTT	AACTGA	GAGACAGTATATGGAATTAAG	4620
QY	4621	AAATTTAG	TGTAAT	GATTTCC	AAAGTGGCT	GCACGACCA	ATGCTGGCTATGCAATGCT	4680


```
Qy 6841 CAGTGCATTTCTGGATCTCTTATCAAAATAGGGAAGTAATCAGCCCATCTGTGTTTGTAAAC 6900
Db 6841 CAGTGCATTTCTGGATCTCTTATCAAAATAGGGAAGTAATCAGCCCATCTGTGTTTGTAAAC 6900
Qy 6901 TATTGACACCTTCAGAGATGGACTCTTTTCGAGCCCATGGAAGTGATGTTGAGCTTGAA 6960
Db 6901 TATTGACACCTTCAGAGATGGACTCTTTTCGAGCCCATGGAAGTGATGTTGAGCTTGAA 6960
Qy 6961 TGGGAAGAAACCACTATGTAATATTCAGACCAAGGATACAAATGGAAGAGATTTTAA 7020
Db 6961 TGGGAAGAAACCACTATGTAATATTCAGACCAAGGATACAAATGGAAGAGATTTTAA 7020
Qy 7021 ATCCAGGGGCGAAGTTACCCCTCATTTCTTCGAAATGGAATGCGAACCTTAAAGAA 7080
Db 7021 ATCCAGGGGCGAAGTTACCCCTCATTTCTTCGAAATGGAATGCGAACCTTAAAGAA 7080
Qy 7081 ATATCTATGCTCTCTCACTGTCCTTT 7108
Db 7081 ATATCTATGCTCTCTCACTGTCCTTT 7108

RESULT 2
AX751409
LOCUS AX751409 6897 bp DNA linear PAT 20-JUN-2003
DEFINITION Sequence 1 from Patent WO03033688.
ACCESSION AX751409
VERSION AX751409.1 GI:32133732
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Liou, J.R.
TITLE Regulation of human receptor tyrosine phosphatase
JOURNAL Patent: WO 03033688-A 1 24-APR-2003;
Bayer Aktiengesellschaft (DE)
FEATURES
    source
        1..6897
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match 95.7%; Score 6800.8; DB 6; Length 6897;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 689; Conservative 0; Mismatches 12; Indels 30; Gaps 2;

Qy 106 ATGGATTTTCTTATCAATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGAT 165
Db 1 ATGGATTTTCTTATCAATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGAT 60
Qy 166 GTTTCGAATGCTGTTCTGTACTAGTAGATATAACCATCTCTTCAATTTCTACAACA 225
Db 61 GTTTCGAATGCTGTTCTGTACTAGTAGATATAACCATCTCTTCAATTTCTACAACA 120
Qy 226 TACACCTCACCTGTACTAGATAGTGACACCAAAATGTAACAAACAGGGGCTCCAGTC 285
Db 121 TACACCTCACCTGTACTAGATAGTGACACCAAAATGTAACAAACAGGGGCTCCAGTC 180
Qy 286 TTCTTAGCCGGGGAAGAGTCGGATCTGCTGGGATTTCTGCTTTGGAATACACCACT 345
Db 181 TTCTTAGCCGGGGAAGAGTCGGATCTGCTGGGATTTCTGCTTTGGAATACACCACT 240
Qy 346 AATCCAAATGGAAGGATATATCTTACATTTGTCGAATATAGGAAGTTGTCGGTGATG 405
Db 241 AATCCAAATGGAAGGATATATCTTACATTTGTCGAATATAGGAAGTTGTCGGTGATG 300
Qy 406 CAAACAGTATATACAAAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAT 465
Db 301 CAAACAGTATATACAAAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAT 360
```

```
Qy 466 CTTAATCTCGGAACAACATATGAATTAAGGTTGCTGCTGAAAAACAGTGTGCGATTGGA 525
Db 361 CTTAATCTCGGAACAACATATGAATTAAGGTTGCTGCTGAAAAACAGTGTGCGATTGGA 420
Qy 526 GTGTTTAGTGATCCATTTCTCTTCCAAACATGCAAGAGTCTCCAGGAAAGTGGTCAAT 585
Db 421 GTGTTTAGTGATCCATTTCTCTTCCAAACATGCAAGAGTCTCCAGGAAAGTGGTCAAT 480
Qy 586 CTCACAGTTGAGCCCTACAAACGCTTCAGCAGTTTAAGCTGATTTTGGTATTTACCTCGCAA 645
Db 481 CTCACAGTTGAGCCCTACAAACGCTTCAGCAGTTTAAGCTGATTTTGGTATTTACCTCGCAA 540
Qy 646 CCAATGCGCAAAATTAACAGCTTCAGAGATAGTGTCAAACATGCGCAGAGTGGATAGTA 705
Db 541 CCAATGCGCAAAATTAACAGCTTCAGAGATAGTGTCAAACATGCGCAGAGTGGATAGTA 600
Qy 706 GTCAAAAGATGCTCAATCAGAGTAGAGGACATTTTTCAGCTGGGAAATTTGCCAGAAATGCAAT 765
Db 601 GTCAAAAGATGCTCAATCAGAGTAGAGGACATTTTTCAGCTGGGAAATTTGCCAGAAATGCAAT 660
Qy 766 GAGAAATAGTGAATCTTTTATGAGGTACAGCCAGCCCTTCTCCAAACCTTTGGTAGATT 825
Db 661 GAGAAATAGTGAATCTTTTATGAGGTACAGCCAGCCCTTCTCCAAACCTTTGGTAGATT 720
Qy 826 ACACCTCCATCGCGTACCAACATTCATCAAGCAGCTTGAACAGAAATGAGATCAGCTCT 885
Db 721 ACACCTCCATCGCGTACCAACATTCATCAAGCAGCTTGAACAGAAATGAGATCAGCTCT 780
Qy 886 GTGTGAAAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTATACAAATATCTT 945
Db 781 GTGTGAAAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTATACAAATATCTT 840
Qy 946 TTTGAAAGTTTCAGCTGCTACACTGAAGCAGGTTTATTTGATAGTAGTACGATTGTGAGAACA 1005
Db 841 TTTGAAAGTTTCAGCTGCTACACTGAAGCAGGTTTATTTGATAGTAGTACGATTGTGAGAACA 900
Qy 1006 CAGAAATCAGTGCCTGAAGACCAACCACTGCTGTTAAGCAGGTAACAGGAAACATCAGAGAAAG 1065
Db 901 CCAGAAATCAGTGCCTGAAGACCAACCACTGCTGTTAAGCAGGTAACAGGAAACATCAGAGAAAG 960
Qy 1066 TCCTTTTCAATTTTATGGGACCCCACTATAGTAAACAGGAAATTTAGTTTATAGATT 1125
Db 961 TCCTTTTCAATTTTATGGGACCCCACTATAGTAAACAGGAAATTTAGTTTATAGATT 1020
Qy 1126 GAATATATGGAACCATCAGTTCGCAATTTTGGATAACAGCAAAAGACCTTCAAGTTTGCA 1185
Db 1021 GAATATATGGAACCATCAGTTCGCAATTTTGGATAACAGCAAAAGACCTTCAAGTTTGCA 1080
Qy 1186 TTCACCTAACCTAACCACTTACAATGATGATGCTATATTGCGGCTGAAACCAAGTGCA 1245
Db 1081 TTCACCTAACCTAACCACTTACAATGATGATGCTATATTGCGGCTGAAACCAAGTGCA 1140
Qy 1246 GGGACTGGGCGCCCAAGTCAAAATTTTCAGTATTTCACTCCACAGATGTTCCAGGGGCGAGTG 1305
Db 1141 GGGACTGGGCGCCCAAGTCAAAATTTTCAGTATTTCACTCCACAGATGTTCCAGGGGCGAGTG 1200
Qy 1306 TTTGATTTTCAACTTCGAGAGGTAGAAATCCACCAAGTAAAGATTTACTTGGAAAGAACCA 1365
Db 1201 TTTGATTTTCAACTTCGAGAGGTAGAAATCCACCAAGTAAAGATTTACTTGGAAAGAACCA 1260
Qy 1366 CGACAAACCAAAATGGAATTTAATCAATACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGA 1425
Db 1261 CGACAAACCAAAATGGAATTTAATCAATACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGA 1320
Qy 1426 ATAAATTTTGGAAATTTTGTCTCACTGGAAATTAATGAGTATATTAATGACCCCTGGCT 1485
Db 1321 ATAAATTTTGGAAATTTTGTCTCACTGGAAATTAATGAGTATATTAATGACCCCTGGCT 1380
Qy 1486 CCAGAAATTTGTGAACATAGTAGAGCAATAGTGGATATATAGAGGTTTCCAGCAGAGATG 1545
Db 1381 CCAGAAATTTGTGAACATAGTAGAGCAATAGTGGATATATAGAGGTTTCCAGCAGAGATG 1440
Qy 1546 TCGTCTGACCTTCACTCACTTGTCTACATTTATATATAACAGCCATCCAGATAAAACTTTT 1605
```

|||||
1441 TGGTCTGACCTTCACTTCTGCTACATTTATATAACAGCCATCCAGATAAAACCTTT 1500
|||||
1606 CTTGCAAGGAATAGAGCTGAAGACCAAGACTTCAACGAGTGTGTAACATAAAGGAATCAGTAT 1665
|||||
1501 CTTGCAAGGAATAGAGCTGAAGACCAAGACTTCAACGAGTGTGTAACATAAAGGAATCAGTAT 1560
|||||
1666 ATTACTGACATTCGACGCTGAACAGCTGTCTTATGTTATCAGGAGACTGTGACCTTTCACT 1725
|||||
1561 ATTACTGACATTCGACGCTGAACAGCTGTCTTATGTTATCAGGAGACTGTGACCTTTCACT 1620
|||||
1726 GAGCACATGATTAGTGTATCTGCTTTCAACCATCATGGGAGAAGGACCAACACAGTTCTC 1785
|||||
1621 GAGCACATGATTAGTGTATCTGCTTTCAACCATCATGGGAGAAGGACCAACACAGTTCTC 1680
|||||
1786 AGTGTTAGGACAGCTCAGCGAAGTGGCAAGCTCAATTAATAAATTAATAACTATAAATAATTT 1845
|||||
1681 AGTGTTAGGACAGCTCAGCGAAGTGGCAAGCTCAATTAATAAATTAATAACTATAAATAATTT 1740
|||||
1846 AGTTCCTTCATCTATTTTGTATATGCGATCCTCCAGAATATCCCAATGGAAAAATAACT 1905
|||||
1741 AGTTCCTTCATCTATTTTGTATATGCGATCCTCCAGAATATCCCAATGGAAAAATAACT 1800
|||||
1906 CACTATACGATTTATGCAATGGAAATGGGATACAAACAGAGCAATCCAGATAACTACCAT 1965
|||||
1801 CACTATACGATTTATGCAATGGAAATGGGATACAAACAGAGCAATCCAGATAACTACCAT 1860
|||||
1966 GATAACAGCTTTCTCATACAGGGTTAAAGAAATACAAAATACAAAATAGAGTGGCA 2025
|||||
1861 GATAACAGCTTTCTCATACAGGGTTAAAGAAATACAAAATACAAAATAGAGTGGCA 1920
|||||
2026 GCCTCAACCCAGATCGGAGAAAGTCTTGTCTCGAAGAAATGACATCTTTGTGAAACT 2085
|||||
1921 GCCTCAACCCAGATCGGAGAAAGTCTTGTCTCGAAGAAATGACATCTTTGTGAAACT 1980
|||||
2086 TCAGAGAGATGAACCGGAATCATCACTCAAGATGTGCGAAGTAAATGATGTACCGCAGAT 2145
|||||
1981 TCAGAGAGATGAACCGGAATCATCACTCAAGATGTGCGAAGTAAATGATGTACCGCAGAT 2040
|||||
2146 GAAATAGGTTGAAGTGGTGCACACCGAAGAACCCCAATGGGATCAATATGCTTATGAA 2205
|||||
2041 GAAATAGGTTGAAGTGGTGCACACCGAAGAACCCCAATGGGATCAATATGCTTATGAA 2100
|||||
2206 GTGCTATATAAATAATAGATCTTTATATATGAAGAACACATCAACACAGACATAATA 2265
|||||
2101 GTGCTATATAAATAATAGATCTTTATATATGAAGAACACATCAACACAGACATAATA 2160
|||||
2266 TTAAGGAACTTAAGACCTCAACCCCTCTATAACAATTTCTGTAAGGCTTTACACAGATTT 2325
|||||
2161 TTAAGGAACTTAAGACCTCAACCCCTCTATAACAATTTCTGTAAGGCTTTACACAGATTT 2220
|||||
2326 GGTCAATGGCAATCAGGTATCTTTTATCTCTGTGAAGGACTTCGGAGACTGTGCTGTAT 2385
|||||
2221 GGTCAATGGCAATCAGGTATCTTTTATCTCTGTGAAGGACTTCGGAGACTGTGCTGTAT 2280
|||||
2386 AGTGCCACCGAATAATACATTACAAAATATTTCTCTGGAGAGATTCGACTATCATTC 2445
|||||
2281 AGTGCCACCGAATAATACATTACAAAATATTTCTCTGGAGAGATTCGACTATCATTC 2340
|||||
2446 CTTTCCCGCAAGTGTCCCAATGGAATCATAAAAAATAATACAAATTTATCTCAAGAGAAAT 2505
|||||
2341 CTTTCCCGCAAGTGTCCCAATGGAATCATAAAAAATAATACAAATTTATCTCAAGAGAAAT 2400
|||||
2506 AATGGAAATAGGAAAGAACTATAATAACAACCTCTTTAAACCAAAACATTAAGTACTG 2565
|||||
2401 AATGGAAATAGGAAAGAACTATAATAACAACCTCTTTAAACCAAAACATTAAGTACTG 2460
|||||
2566 AAGAAATATACCAATATATCATTTGAGGTCTGTAGTACATCGAAGGTGAAGAGTT 2625
|||||
2461 AAGAAATATACCAATATATCATTTGAGGTCTGTAGTACATCGAAGGTGAAGAGTT 2520
|||||
2626 CGGAGTCTCCCATAGTATACCTGACGGAGGAAGATGCTCCTGATTTCTCCCCCTCAAGAC 2685
|||||

Db
2521 CGGAGTGTCTCCATAGTATATACTGACGGAGGAAGATGCTCTGATTTCTCCCTCTCAAGAC 2580
QY
2686 TTCTCTGTAAACACAGTTGCTGTGTGTCACGGTGAAGTGTGTCATGCGCAACACCCCTGGAG 2745
Db
2581 TTCTCTGTAAACACAGTTGCTGTGTGTCACGGTGAAGTGTGTCATGCGCAACACCCCTGGAG 2640
QY
2746 CCAATAGGAATTTATCTTTATTTACACAGTTTATGCTGCGAATAGATCATCATTAATAAAT 2805
Db
2641 CCAATAGGAATTTATCTTTATTTACACAGTTTATGCTGCGAATAGATCATCATTAATAAAT 2700
QY
2806 ATTAATGTCTCACTGAAACATCATTTGGAGTTATCAGATTTGGATTAATGTTGAAATACAGT 2865
Db
2701 ATTAATGTCTCACTGAAACATCATTTGGAGTTATCAGATTTGGATTAATGTTGAAATACAGT 2760
QY
2866 GCTTATGTAACAGCTAGCACAGATTTGGTGTGTAAGGAAACAGAGCAATATCATTAGC 2925
Db
2761 GCTTATGTAACAGCTAGCACAGATTTGGTGTGTAAGGAAACAGAGCAATATCATTAGC 2820
QY
2926 TTTTCAACACACAGAGGGAGCACCAAGCGATCCTCCCAAGATGTTTATTTATGCAAACTTC 2985
Db
2821 TTTTCAACACACAGAGGGAGCACCAAGCGATCCTCCCAAGATGTTTATTTATGCAAACTTC 2880
QY
2986 AGTTCCTTCATCAATTAATTTCTTTCTGGACACCTCCTTCAAACTTAATGGGATTTACAA 3045
Db
2881 AGTTCCTTCATCAATTAATTTCTTTCTGGACACCTCCTTCAAACTTAATGGGATTTACAA 2940
QY
3046 TATTACTCTGTTTATTTACAGAAATACTTCAAGTACTTTTATGCAGAAATTTTACACTCCAT 3105
Db
2941 TATTACTCTGTTTATTTACAGAAATACTTCAAGTACTTTTATGCAGAAATTTTACACTCCAT 3000
QY
3106 GAACTAAACCAATGACTTTTGACAAATATGCTGTATCCACAATTTATAGATAAACTGCAATA 3165
Db
3001 GAAGTAAACCAATGACTTTTGACAAATATGCTGTATCCACAATTTATAGATAAACTGCAATA 3060
QY
3166 TTCAGTACTATAFACATTTTGGTTAAACAGCAAGTACTTCAAGTGGAAATGGGAATAAAGC 3225
Db
3061 TTCAGTACTATAFACATTTTGGTTAAACAGCAAGTACTTCAAGTGGAAATGGGAATAAAGC 3120
QY
3226 AGTGACATCATTTGAAAGTATACACAGATCAAGACATACCTGAAAGGTTTGTGAAACCTTG 3285
Db
3121 AGTGACATCATTTGAAAGTATACACAGATCAAGACATACCTGAAAGGTTTGTGAAACCTTG 3180
QY
3286 ACTTACGAATCCATTTCTGTCAACTGCAATAAATGTAAGCTGGGTCCCAACGGCTCAACCA 3345
Db
3181 ACTTACGAATCCATTTCTGTCAACTGCAATAAATGTAAGCTGGGTCCCAACGGCTCAACCA 3240
QY
3346 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGAGACTCCTCGCCATGTGAGA 3405
Db
3241 AACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGAGACTCCTCGCCATGTGAGA 3300
QY
3406 CCACCTCTGTTTACATATCAGAGAGACATATATTTTGAATACTGGAATAAATACACTGAT 3465
Db
3301 CCACCTCTGTTTACATATCAGAGAGACATATATTTTGAATACTGGAATAAATACACTGAT 3360
QY
3466 TATATAATTAATAATTAATCTCCATCAACAGAAAAAGGATTTCTTGATACTTATCTGCCAG 3525
Db
3361 TATATAATTAATAATTAATCTCCATCAACAGAAAAAGGATTTCTTGATACTTATCTGCCAG 3420
QY
3526 CTATACATCAAGACTGGAAGAGATGTCCAGAAAACTTCAACCAATTAATCAACACTTTTAA 3585
Db
3421 CTATACATCAAGACTGGAAGAGATGTCCAGAAAACTTCAACCAATTAATCAACACTTTTAA 3480
QY
3586 AACCTTTCTCTACCTCAGTCTCTTATCATGCGATCCCGCAAGTAAAGCAATGTCGA 3645
Db
3481 AACCTTTCTCTACCTCAGTCTCTTATCATGGATCCCGCAAGTAAAGCAATGTCGA 3540
QY
3646 ATAATAAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAAATTTATTTTCACTACTTCT 3705
Db
3541 ATAATAAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAAATTTATTTTCACTACTTCT 3600
QY
3706 GATAATTAATTAATGGAAGACTTTCACATTTTACATTTATAGCTTTTGTGCTGCC 3765
Db
3601 GATAATTAATTAATGGAAGACTTTCACCAATTTACATTTATAGCTTTTGTGCTGCC 3660


```
|||||
Db 5821 TCCTCTCAGATGACAGAAATTAATGACACTAAATGGAAGTCGATCAGTCATCAGATG 5880
QY GCAGACTGGAACCTGAAGACGAGAGATTAACCGC-----
Db 5881 GCAGACTGGAACCTGAAGACGAGAGATTAACCGCGTTACTTTAGTTATAGAAAATCCATC 5940
QY 6021 --GCCAATAGCAAGAAATCTCTCCGCAACATGTTGAGAGCTTTGCAACACACAC 6078
Db 5941 AAGCCAAATAGCAAGAAATCTCTCCGCAACATGTTGAGAGCTTTGCAACACACAC 6000
QY 6079 CTAAGGTTCAAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTCTTCAACT 6138
Db 6001 CTAAGGTTCAAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTCTTCAACT 6060
QY 6139 GATGCTGATCTGCTTGGATAGAGCAAAACCGTTTCCCAAAATATAAACAATATAAT 6198
Db 6061 GATGCTGATCTGCTTGGATAGAGCAAAACCGTTTCCCAAAATATAAACAATATAAT 6117
QY 6199 AATAATAACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATATAT 6258
Db 6118 AATAATAACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGGTTCCGATTTATATAT 6177
QY 6259 GCCAGCTATATTTCTGTTATTTATGTCCAAAATGAATTTATGCTACTCAAGTCCACTA 6318
Db 6178 GCCAGCTATATTTCTGTTATTTATGTCCAAAATGAATTTATGCTACTCAAGTCCACTA 6237
QY 6319 CCAAGAACAGTTCGAGATTTTGGAGATTTGGAGATTTGGAGAACCAAGGCAAAACATTA 6378
Db 6238 CCAAGAACAGTTCGAGATTTTGGAGATTTTGGAGATTTTGGAGAACCAAGGCAAAACATTA 6297
QY 6379 ATGCTAACACAGTGTGTTGAAAAGGACGATCAGATGCCATGATTTGGCCAGAGGAC 6438
Db 6298 ATGCTAACACAGTGTGTTGAAAAGGACGATCAGATGCCATGATTTGGCCAGAGGAC 6357
QY 6439 AACAAAGCAGTACTGCTTTGGAGATATAGTGAATACAAAGCTAATGAGGATGTTCAA 6498
Db 6358 AACAAAGCAGTACTGCTTTGGAGATATAGTGAATACAAAGCTAATGAGGATGTTCAA 6417
QY 6499 ATAGATGGAATATACAGGATCTGAAAATTTGAAAGCATGAGGATGCAATGATTCGA 6558
Db 6418 ATAGATGGAATATACAGGATCTGAAAATTTGAAAGCATGAGGATGCAATGATTCGA 6477
QY 6559 CAGGTAACTTTACTGCTGCGCAGAGCATGAGGTTCTCGAGAACAGCGCCCTCTAAT 6618
Db 6478 CAGGTAACTTTACTGCTGCGCAGAGCATGAGGTTCTCGAGAACAGCGCCCTCTAAT 6537
QY 6619 CACTTTGTGAAGTTGGTTCGAGCAAGCAGGCAATGACACACACACTATGATGTTTCA 6678
Db 6538 CACTTTGTGAAGTTGGTTCGAGCAAGCAGGCAATGACACACACTATGATGTTTCA 6597
QY 6679 TGCACTGCTGGAGTTGGAGAACTGAGATTTTATGCTCTGGAACATTTAAACAACAT 6738
Db 6598 TGCACTGCTGGAGTTGGAGAACTGAGATTTTATGCTCTGGAACATTTAAACAACAT 6657
QY 6739 ATAAATGACCAATTTTGGATATATATGGAATAGTACTAGTGAATCTGAGAGTGAAGA 6798
Db 6658 ATAAATGACCAATTTTGGATATATATGGAATAGTACTAGTGAATCTGAGAGTGAAGA 6717
QY 6799 ATGTCATGTCGAGAACTGCGACAGATATCTTTTACACAGTGCAATCTGGAATCTC 6858
Db 6718 ATGTCATGTCGAGAACTGCGACAGATATCTTTTACACAGTGCAATCTGGAATCTC 6777
QY 6859 TTAATAAATAGGGAATATACAGCCATCTGTTTGTGTTAACTAATTCAGCACTTCAGAAG 6918
Db 6778 TTAATAAATAGGGAATATACAGCCATCTGTTTGTGTTAACTAATTCAGCACTTCAGAAG 6837
QY 6919 ATGCACTTTGCAAGCCATGGAAGTGTGTTGAGCTTGAATGGAAGAAACCACTATG 6978
Db 6838 ATGCACTTTGCAAGCCATGGAAGTGTGTTGAGCTTGAATGGAAGAAACCACTATG 6897
```

```
AX644236
LOCUS AX644236 7225 bp DNA linear PAT 27-FEB-2003
DEFINITION Sequence 10 from Patent WO02090530.
ACCESSION AX644236
VERSION AX644236.1 GI:28610295
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Lee, E.A., Wallia, N.K., Baughn, M.R., Ison, C.H., Gururajan, R.,
Arvizu, C., Yao, M.G., Jackson, J.L., Tang, T.Y., Yue, H., Tran, B.,
Ding, L., Lu, D.A., Lal, P.G. and Warren, B.A.
Kinases and phosphatases
Patent: WO 02090530-A 10 14-NOV-2002;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..7225
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7480588CB1"
ORIGIN
Query Match 95.7%; Score 6799.8; DB 6; Length 7225;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 6962; Conservative 0; Mismatches 17; Indels 72; Gaps 8;
QY 106 ATGGATTTTCTTATCATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGAT 165
DB 1 ATGGATTTTCTTATCATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGAT 60
QY 166 GTTTCCAATGTCGTTCTGTTACTAGGTACGATATACCATCTCTTCAATTTCTACACA 225
DB 61 GTTTCCAATGTCGTTCTGTTACTAGGTACGATATACCATCTCTTCAATTTCTACACA 120
QY 226 TACACTCACCCTGTTACTAGATAGTACACCAATATGTAACAAACACAGGCTCCAGTC 285
DB 121 TACACTCACCCTGTTACTAGATAGTACACCAATATGTAACAAACACAGGCTCCAGTC 165
QY 286 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTGTTGGAATACACCACT 345
DB 166 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTGTTGGAATACACCACT 225
QY 346 AATCCAAATGGAAGATTTATCTTACATTTGTCAAATATAAGAAAGTTGTCGGTGAATG 405
DB 226 AATCCAAATGGAAGATTTATCTTACATTTGTCAAATATAAGAAAGTTGTCGGTGAATG 285
QY 406 CAAACAGTATATACACAGTCAGATCAAGCCAGACAGTCGGAAGTTCTTCTTACTAAT 465
DB 286 CAAACAGTATATACACAGTCAGATCAAGCCAGACAGTCGGAAGTTCTTCTTACTAAT 345
QY 466 CTTAATCTCTGGAACAACTATGAAATTAAGTTGCTGCTGGAACAACTGCTGGAATGGA 525
DB 346 CTTAATCTCTGGAACAACTATGAAATTAAGTTGCTGCTGGAACAACTGCTGGAATGGA 405
QY 526 GTGTTTAGTGAATCTTCTTCCAACTGCAAGAAAGTCTCCAGGAAAGTGGTGAAT 585
DB 406 GTGTTTAGTGAATCTTCTTCCAACTGCAAGAAAGTCTCCAGGAAAGTGGTGAAT 465
QY 586 CTCACAGTTGAGCCCTACAACTGTTTACAGCTGATTTGTTTACTCTCGGCA 645
DB 466 CTCACAGTTGAGCCCTACAACTGTTTACAGCTGATTTGTTTACTCTCGGCA 525
QY 646 CCAATAGGCAAAATTAACAGCTTCAAGATTTAGTGTCAACATGCAAGAGTGGGATGTA 705
DB 526 CCAATAGGCAAAATTAACAGCTTCAAGATTTAGTGTCAAGATGCAAGAGTGGGATGTA 585
QY 706 GTGAAAGATGTTCAATCAGAGTAGAGCAATTTTGAATGAGGAAATGCGAGAAATGCAAT 765
DB 586 GTGAAAGATGTTCAATCAGAGTAGAGCAATTTTGAATGAGGAAATGCGAGAAATGCAAT 645
```

QY 766 GAGGATAGTGAATCTTTTATGAGGTACAGCGAGCCCTTCTCCAGCCCTTGAGAGTT 825
Db 646 GAGGATAGTGAATCTTTTATGAGGTACAGCGAGCCCTTCTCCAGCCCTTGAGAGTT 705
QY 826 ACACCTCCATCGGTGACACACATTCATCAAGACAGCTTGACACAGATGAGATCAGCTCT 885
Db 706 ACACCTCCATCGGTGACACACATTCATCAAGACAGCTTGACACAGATGAGATCAGCTCT 765
QY 886 GTGTGGAAGAGGCTATCAGTTTGTAGTGACACACCTTGAGACCTTATACAAATATCTT 945
Db 766 GTGTGGAAGAGGCTATCAGTTTGTAGTGACACACCTTGAGACCTTATACAAATATCTT 825
QY 946 TTTGAGATTTTCACTGCTGCTACAACTGAGCAGGTTTATTTGATAGTACGATGTCAGAAC 1005
Db 826 TTTGAGATTTTCACTGCTGCTACAACTGAGCAGGTTTATTTGATAGTACGATGTCAGAAC 885
QY 1006 CCAGATCAGTGTGAGGACACACCAAACTGCTGTAACAGGCAACATCAGAGAAAG 1065
Db 886 CCAGATCAGTGTGAGGACACACCAAACTGCTGTAACAGGCAACATCAGAGAAAG 945
QY 1066 TCCTTTTCAATTTTATGGGACCCCACTATAGTAAACAGGGAATTTAGTTATAGAGTT 1125
Db 946 TCCTTTTCAATTTTATGGGACCCCACTATAGTAAACAGGGAATTTAGTTATAGAGTT 1005
QY 1126 GAATTTATAGGACAT-----CAGTGCATTTTGGATTAACAGCAAAAGACCTCAAG 1179
Db 1006 GAATTTATAGGACATCAGGTGCGAGTGCATTTTGGATTAACAGCAAAAGACCTCAAG 1065
QY 1180 TTTGCACTTCACTAACCTTAACACATTTACAACTGATGATGCTATATTGGGCTGAAACC 1239
Db 1066 TTTGCACTTCACTAACCTTAACACATTTACAACTGATGATGCTATATTGGGCTGAAACC 1125
QY 1240 AGTGCAGGGACTGGGCCCCAAGTCAAAATTTTCACTGATTTCACTCCACAGATGTCAGGG 1299
Db 1126 AGTGCAGGGACTGGGCCCCAAGTCAAAATTTTCACTGATTTCACTCCACAGATGTCAGGG 1185
QY 1300 GCAGTGTGTTGATTTACACTTGAGAGGTAGATTCACGCAAGTAAGATTAAGTGAAG 1359
Db 1186 GCAGTGTGTTGATTTACACTTGAGAGGTAGATTCACGCAAGTAAGATTAAGTGAAG 1245
QY 1360 AAACCAACGACAAACAAATGGAATTTTAAACCAATACGAGTGAAGTGTAGTTCAGAG 1419
Db 1246 AAACCAACGACAAACAAATGGAATTTTAAACCAATACGAGTGAAGTGTAGTTCAGAG 1305
QY 1420 ACAGGAATAATTTTGGAAATACTTTGCTCACTGGAATAATCAGTATATAAATGACCCC 1479
Db 1306 ACAGGAATAATTTTGGAAATACTTTGCTCACTGGAATAATGAG--ATAAATGACCCC 1362
QY 1480 ATGGCTCCAGAAATTTGTGAACATAGTAGAGCCAAATGGTAGGATTTATAGGGGTTGAG 1539
Db 1363 ATGGCTCCAGAAATTTGTGAACATAGTAGAGCCAAATGGTAGGATTTATAGGGGTTGAG 1422
QY 1540 GAGATGCTGCTGACCTTCACTCACTGCTGACATTTATATATAACAGCCATCCAGATAA 1599
Db 1423 GAGATGCTGCTGACCTTCACTCACTGCTGACATTTATATATAACAGCCATCCAGATAA 1482
QY 1600 AACTTTCTGCAAGGAATAGAGCTGGAAGCCAGACTTCACTGCTGTAACCTCAAGGAAT 1659
Db 1483 AACTTTCTGCAAGGAATAGAGCTGGAAGCCAGACTTCACTGCTGTAACCTCAAGGAAT 1542
QY 1660 CAGTATATTAAGTGAACATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1719
Db 1543 CAGTATATTAAGTGAACATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1602
QY 1720 TTTCACTGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1779
Db 1603 TTTCACTGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1662
QY 1780 GTTCTCAGTGTAGGACAGCTCAGCAAGTGCAGAGCTTCAATTAATAATTAATAATAA 1839
Db 1663 GTTCTCAGTGTAGGACAGCTCAGCAAGTGCAGAGCTTCAATTAATAATTAATAATAA 1722
QY 1840 AATATTAGTCTTCACTATTTTGTATATTTGGATCTCTCCAGAAATCCCAATGGAAA 1899

Db 1723 AATATTAGTCTTCACTATTTTGTATATTTGGATCTCTCCAGAAATCCCAATGGAAA 1782
QY 1900 ATAACCTCACTATACGATTTTATGCAATTTGGAATTTGGAATCAAAACAGAGCAATTTCCAGATAACT 1959
Db 1783 ATAACCTCACTATACGATTTTATGCAATTTGGAATTTGGAATCAAAACAGAGCAATTTCCAGATAACT 1842
QY 1960 ACCATAGATAACAGCTTTTCTCATACAG-----GGTTAAAGAAATACACAAA 2007
Db 1843 ACCATAGATAACAGCTTTTCTCATACAGGTAGAAAACAAATGGTTAAAGAAATACACAAA 1902
QY 2008 TACAAAATCAGAGTGCAGCCTCAACCCACAGATGGAAGAAAGTTCTTTGCTGGAAGAAAT 2067
Db 1903 TACAAAATCAGAGTGCAGCCTCAACCCACAGTTCAGAAAGTTCTTTGCTGGAAGAAAT 1962
QY 2068 GACATCTTTGTGAGAACTTTCAGAAATGAAACCGGAATCATCACCTCAAGATGTCGAAGTA 2127
Db 1963 GACATCTTTGTGAGAACTTTCAGAAATGAAACCGGAATCATCACCTCAAGATGTCGAAGTA 2022
QY 2128 ATTGATGTTTCCGAGATGAAATAAGTTGAAAGTGTCAACCCGAAAGCCCAATGGG 2187
Db 2023 ATTGATGTTTCCGAGATGAAATAAGTTGAAAGTGTCAACCCGAAAGCCCAATGGG 2082
QY 2188 ATCATTTATTTGCTTATGAGTGTCTATATAAATATATAGTACTTTATATATGAGAACACA 2247
Db 2083 ATCATTTATTTGCTTATGAGTGTCTATATAAATATATAGTACTTTATATATGAGAACACA 2142
QY 2248 TCAACAAACAGACATATAATTTAAGGAACTTAAGACCTTCAACCTCTATAAATTTCTGTA 2307
Db 2143 TCAACAAACAGACATATAATTTAAGGAACTTAAGACCTTCAACCTCTATAAATTTCTGTA 2202
QY 2308 AGGTCTTACACAGATTTGGTCAATGAGTATCTTTTACTCTCTGTAAGGACT 2367
Db 2203 AGGTCTTACACAGATTTGGTCAATGAGTATCTTTTACTCTCTGTAAGGACT 2262
QY 2368 TGGGAGACTGCTGCTGATAGTGCACAGAAATATACATTAACAAATATTTCTTCTGGA 2427
Db 2263 TGGGAGACTGCTGCTGATAGTGCACAGAAATATACATTAACAAATATTTCTTCTGGA 2322
QY 2428 GAGATTGACTATCACTTCTTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACA 2487
Db 2323 GAGATTGACTATCACTTCTTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACA 2382
QY 2488 ATTTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATATAACAACTCTTTTAAACC 2547
Db 2383 ATTTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATATAACAACTCTTTTAAACC 2442
QY 2548 CAAACATTTAAAGTACTGAGAAATATACCCAAATATATCAATTTGAGGTGCTGCTAGTACA 2607
Db 2443 CAAACATTTAAAGTACTGAGAAATATACCCAAATATATCAATTTGAGGTGCTGCTAGTACA 2502
QY 2608 CTGAAAGGTGAAGGAGTTCGGAGTGTCTCCATAAGTATACTGACGAGGAAGATGCTCCT 2667
Db 2503 CTGAAAGGTGAAGGAGTTCGGAGTGTCTCCATAAGTATACTGACGAGGAAGATGCTCCT 2562
QY 2668 GATTCTCCCCCTCAAGACTTCTCTGTAAACAGTTGTCTGGTGTCAACGTTGATGTC 2727
Db 2563 GATTCTCCCCCTCAAGACTTCTCTGTAAACAGTTGTCTGGTGTCAACGTTGATGTC 2622
QY 2728 TGGCAACCAACCTGAGGCAAAATGGAATTTCTTTATTTATACAGTTTATGTCT---GG 2784
Db 2623 TGGCAACCAACCTGAGGCAAAATGGAATTTCTTTATTTATACAGTTTATGTCTGGAGG 2682
QY 2785 AATAGATCATCATTTAAATACTTAAATGTCACTGAAAACATCAATGGAGTTTATCAGATTTG 2844
Db 2683 AATAGATCATCATTTAAATACTTAAATGTCACTGAAAACATCAATGGAGTTTATCAGATTTG 2742
QY 2845 GATTATAATGTTGAAATACAGTGTCTTATGTAACAGCTAGCACAGATTTTGGTATGGGAAA 2904
Db 2743 GATTATAATGTTGAAATACAGTGTCTTATGTAACAGCTAGCACAGATTTTGGTATGGGAAA 2802
QY 2905 ACAGGAGCAATATCATTTAGCTTTTCAACACAGAGGGAGCACCAAGCGATCTCTCCAAA 2964

QY 5125 ATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCCTACTGCTGCTCCAGATTCC 5184
DB 5020 ATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCCTACTGCTGCTCCAGATTCC 5079
QY 5185 AACCTCAGTATATACAGAAACCAACACATCTTCTGCAATGCTAGAGGACTAAAA 5244
DB 5080 AACCTCAGTATATACAGAAACCAACACATCTTCTGCAATGCTAGAGGACTAAAA 5139
QY 5245 GGTGGACATACATCAATATACAGTGTGTTTACGCAATGCTAGTGTGCTGAGGTCCTCAAG 5304
DB 5140 GGTGGACATACATCAATATACAGTGTGTTTACGCAATGCTAGTGTGCTGAGGTCCTCAAG 5199
QY 5305 GTTCCGATGAGATTAACCATGATATCAAGCTCCAGCAGCAACAAAAACCAACCAACC 5364
DB 5200 GTTCCGATGAGATTAACCATGATATCAAGCTCCAGCAGCAACAAAAACCAACCAACC 5259
QY 5365 CCTATTATGATGTCACAGGAAACTGCTTGTGACTTCAACAAATTAACAATCAGAAATG 5424
DB 5260 CCTATTATGATGTCACAGGAAACTGCTTGTGACTTCAACAAATTAACAATCAGAAATG 5319
QY 5425 CCAATATGTTACTACAGTGATGATGACGACCAATAAAAAATGTACAAATGCTTGGGACA 5484
DB 5320 CCAATATGTTACTACAGTGATGATGACGACCAATAAAAAATGTACAAATGCTTGGGACA 5379
QY 5485 GAAACAGGACTCAGCATGATGGAATGTAAACAAAGTGTATGATGATATTTTAAATAA 5544
DB 5380 GAAACAGGACTCAGCATGATGGAATGTAAACAAAGTGTATGATGATATTTTAAATAA 5439
QY 5545 GCAAGGCCATATTTTACAATGAAGGCTTTTCTTAACCTCCATGTAACAGAAAGAAAGACA 5604
DB 5440 GCAAGGCCATATTTTACAATGAAGGCTTTTCTTAACCTCCATGTAACAGAAAGAAAGACA 5499
QY 5605 AAGTTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGATTAATGATGCATGATTCCT 5664
DB 5500 AAGTTTAGTGGCAATGAAGAAATCTACATCATAGGTGCTGATTAATGATGCATGATTCCT 5559
QY 5665 GGCATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAAGCAATCTTATTTAA 5724
DB 5560 GGCATGAAGACAAAATTTGCAATGGACCACTGAAACCAAAAAAGCAATCTTATTTAA 5619
QY 5725 TTTAGAGTACAATATTTATGGGACAAATTTACTGACTCTGATTAATCTGACCTGTTAAG 5784
DB 5620 TTTAGAGTACAATATTTATGGGACAAATTTACTGACTCTGATTAATCTGACCTGTTAAG 5679
QY 5785 ACTTTAGGGGAAGGACTTTTCAAGAAACCGTAGAGATCAATCTTTCCGTCATCTTGTGT 5844
DB 5680 ACTTTAGGGGAAGGACTTTTCAAGAAACCGTAGAGATCAATCTTTCCGTCATCTTGTGT 5739
QY 5845 ATCCTTTCAATAATTTCTCTTGGAAACAGCTATTTTGGCAATTTGCAATTTGCAAGAA 5904
DB 5740 ATCCTTTCAATAATTTCTCTTGGAAACAGCTATTTTGGCAATTTGCAATTTGCAAGAA 5799
QY 5905 CAGAAAGAGGTGGACATCTCTCTCAGGATGCAAGAAATTTGCAATTTGCAATTTGCAAG 5964
DB 5800 CAGAAAGAGGTGGACATCTCTCTCAGGATGCAAGAAATTTGCAATTTGCAATTTGCAAG 5859
QY 5965 CTGGATCAGCTCATACAGTGGCAGACCTGGAATGGAAGCAGAGATTAACGCG --- 6020
DB 5860 CTGGATCAGCTCATACAGTGGCAGACCTGGAATGGAAGCAGAGATTAACGCGGTTA 5919
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTCTGCAACATGTTGAA 6057
DB 5920 CTTTAGTTATAGAAAATCCATCAAGCCCAATAAGCAAGAAATCTCTCTGCAACATGTTGAA 5979
QY 6058 GAGCTTTGCAAAAACCAACCTTAAGTTTCAAGAAAGAAATTTTGGAAATTAACAAATTT 6117
DB 5980 GAGCTTTGCAAAAACCAACCTTAAGTTTCAAGAAAGAAATTTTGGAAATTTACCAATTT 6039
QY 6118 CTTTCAGATCTTTCTTCACTGATGCTGCTGCTTGGAAATGAGCAAAAACCGCTTC 6177
DB 6040 CTTTCAGATCTTTCTTCACTGATGCTGCTGCTTGGAAATGAGCAAAAACCGCTTC 6099
QY 6178 CCAACATRAAACCATATATATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGT 6237

DB 6100 CCAACATRAAACCATATATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGT 6156
QY 6238 CCAAGTTTCGATATATTAATGCGAGCTATATTTCTGTTTATTTATGTCCTCAATGAAATTT 6297
DB 6157 CCAAGTTTCGATATATTAATGCGAGCTATATTTCTGTTTATTTATGTCCTCAATGAAATTT 6216
QY 6298 ATTGCTACTCAAGGTCACCTACAGGAAACAGTTGGAGATTTTGGAGAAATGTTGTTGGAA 6357
DB 6217 ATTGCTACTCAAGGTCACCTACAGGAAACAGTTGGAGATTTTGGAGAAATGTTGTTGGAA 6276
QY 6358 ACCAGGCGAAACATTTAGTAACTGTAACACAGTGTGTTTGAAGAAAGCGATCAGATGC 6417
DB 6277 ACCAGGCGAAACATTTAGTAACTGTAACACAGTGTGTTTGAAGAAAGCGATCAGATGC 6336
QY 6418 CATCAGTATTTGGCCAGAGGACCAACAGCCAGTTTACTGTTCTTGGAGATATAGTATTAACA 6477
DB 6337 CATCAGTATTTGGCCAGAGGACCAACAGCCAGTTTACTGTTCTTGGAGATATAGTATTAACA 6396
QY 6478 AAGCTAATGGAGATGTTTCAATATAGATGGAATCATCAGGATCTGAAATTTGAAAGGCAAT 6537
DB 6397 AAGCTAATGGAGATGTTTCAATATAGATGGAATCATCAGGATCTGAAATTTGAAAGGCAAT 6456
QY 6538 GGGGATTCGATGACTGTTGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCT 6597
DB 6457 GGGGATTCGATGACTGTTGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCT 6516
QY 6598 GAGAACAGCGCCCTCTAATTTCTGTAAGTGTGTTTCGAGCAAGCGGACATGAC 6657
DB 6517 GAGAACAGCGCCCTCTAATTTCTGTAAGTGTGTTTCGAGCAAGCGGACATGAC 6576
QY 6658 ACCACACTATGATGTTTCACTGCTGAGTGTGGAAGAACTGGAGTTTATTTGCT 6717
DB 6577 ACCACACTATGATGTTTCACTGCTGAGTGTGGAAGAACTGGAGTTTATTTGCT 6636
QY 6718 CTGGACCATTTAACAACAATATTAATGACCATGATTTTGTGGATATATATGACCTAGTA 6777
DB 6637 CTGGACCATTTAACAACAATATTAATGACCATGATTTTGTGGATATATATGACCTAGTA 6696
QY 6778 GCTGAATCAGAAAGTGAAGAAATGTCATGTTGCAAACTGCGACAGATATCTTTTAA 6837
DB 6697 GCTGAATCAGAAAGTGAAGAAATGTCATGTTGCAAACTGCGACAGATATCTTTTAA 6756
QY 6838 CACCACTGATTTCTGGATCTTTATCAAAATAGGGAAGTAATCAGCCCATCTGTTTGT 6897
DB 6757 CACCACTGATTTCTGGATCTTTATCAAAATAGGGAAGTAATCAGCCCATCTGTTTGT 6816
QY 6898 AACTATTACGACTTTCAGAAAGTGGACTCTTTGAGCGCATGGAAGGTGATGTTGAGCTT 6957
DB 6817 AACTATTACGACTTTCAGAAAGTGGACTCTTTGAGCGCATGGAAGGTGATGTTGAGCTT 6876
QY 6958 GAATGGGAAGAAACCACTATGTAATATTCAGACCAAGAAAGTACAAATTCGAGAGATTT 7017
DB 6877 GAATGGGAAGAAACCACTATGTAATATTCAGACCAAGAAAGTACAAATTCGAGAGATTT 6936
QY 7018 TAAATCCAGGGGCCAAAAGTTACCCCTCATTTCTCCGAAATTTGAAATGTCACACCTTAA 7077
DB 6937 TAAATCCAGGGGCCAAAAGTTACCCCTCATTTCTCCGAAATTTGAAATGTCACACCTTAA 6996
QY 7078 GAAATATCTATCTTCTCTCACTGTCCTTT 7108
DB 6997 GAAATATCTATCTTCTCTCACTGTCCTTT 7027

RESULT 4

AX686453
LOCUS AX686453
DEFINITION Sequence 9 from Patent WO20057450.
ACCESSION AX686453
VERSION AX686453.1 GI:29372157
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

AX686453
Sequence 9 from Patent WO20057450.
AX686453
AX686453.1
GI:29372157
Homo sapiens (human)
Homo sapiens

linear
PAT 29-MAR-2003

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
Edinger, S., Macdougall, J.R., Millet, I., Ellerman, K., Stone, D.J.,
Gerlach, V., Grosse, W.M., Alsbrook, J.P., Lepley, D.M., Rieger, D.,
Burgess, C.E., Casman, S.J., Spytek, K.A., Boldog, F.L., Li, L.,
Padigara, M., Mishra, V., Patturajan, M., Shenoy, S., Rastelli, L.,
Tchernev, V.T., Vernet, C.A., Zehusen, B.D., Malyankar, U.M., Guo, Y.,
Miller, C.E. and Gangolli, E.A.
Proteins and nucleic acids encoding same
Patent: WO 02057450-A 9 25-JUL-2002;
Curagen Corporation (US)
Location/Qualifiers
1. .6903
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

FEATURES
Source

ORIGIN

Query Match 94.0%; Score 6683.6; DB 6; Length 6903;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 6842; Conservative 0; Mismatches 19; Indels 57; Gaps 9;

QY 106 ATGGATTTTCTTATCAATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTTGAT 165
DB 1 ATGGATTTTCTTATCAATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTTGAT 60

QY 166 GTTTCGAATGCTGCTCTGCTAGGTACGATATTAACGATCTCTTCAATTTCTACACA 225
DB 61 GTTTCGAATGCTGCTCTGCTAGGTACGATATTAACGATCTCTTCAATTTCTACACA 120

QY 226 TACACCTCACTGTTACTAGAAAGTAGACACCAAAATGTAACAAAACAGGGCTCCAGTC 285
DB 121 TACACCTCACTGTTACTAGAAAGTAGACACCAAAATGTAACAAAACAGGGCTCCAGTC 180

QY 286 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTGTTTGGAAATACACACCT 345
DB 181 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTGTTTGGAAATACACACCT 240

QY 346 AATCCAAATGGAAGGATATATCTTACATTTGTCGAATATGAAGGATTTGTCGGTGGATG 405
DB 241 AATCCAAATGGAAGGATATATCTTACATTTGTCGAATATGAAGGATTTGTCGGTGGATG 300

QY 406 CAAACAGTATATACAGAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 465
DB 301 CAAACAGTATATACAGAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAAT 360

QY 466 CTTTAATCTCGGAACACATATGAATTAAGTTGCTGCTGAAACAGTGTGCGCATTTGGA 525
DB 361 CTTTAATCTCGGAACACATATGAATTAAGTTGCTGCTGAAACAGTGTGCGCATTTGGA 420

QY 526 GTGTTTAGTATCCATTTCTCTTCCAACTGCGAAGT---GCTCCAGGAAAGTGGTG 582
DB 421 GTGTTTAGTATCCATTTCTCTTCCAACTGCGAAGTCCAGCTCCAGGAAAGTGGTG 480

QY 583 AATCTCAGTTCAGGCTTACAGCTTACAGCTTACAGCTTACAGCTTACAGCTTACAGCT 642
DB 481 AATCTCAGTTCAGGCTTACAGCTTACAGCTTACAGCTTACAGCTTACAGCTTACAGCT 540

QY 643 CAACCAATGGCAAAATTTACAGCTTCAAGTTAGTGTCAAAATGCTCCAGAGTGGGATA 702
DB 541 CAACCAATGGCAAAATTTACAGCTTCAAGTTAGTGTCAAGCATGCCAGAGTGGGATA 600

QY 703 GTAGTGAAGATGTCTCAATCAGAGTAGAGCAATTTTGACTGGGAAATTTGCCAGATGC 762
DB 601 GTAGTGAAGATGTCTCAATCAGAGTAGAGCAATTTTGACTGGGAAATTTGCCAGATGC 660

QY 763 AAT---GAGAAATAGTGAATCTTTTATGAGTACAGCCAGCTTCTCCAACTTGGT 819
DB 661 AATGTAGAGAAATAGTGAATCTTTTATGAGTACAGCCAGCTTCTCCAACTTGGT 720

QY 820 AGAGTTACCTCCATCGCTTACCAATTCATCAAGCACTGTGACACAGAAATGAGATC 879
DB 1899 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1899
DB 1738 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1797
QY 1900 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1959
DB 1798 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1857

DB 721 AGAGTTACACCTCCATCCGTCACCAACATTCATCAAGCACTGTGACACAGAAATGAGATC 780
QY 880 AGCTCTGTGTGGAAGAGCCTATCAGTTTGTAGTGAACACACTTGAGACCTTATACACA 939
DB 781 AGCTCTGTGTGGAAGAGCCTATCAGTTTGTAGTGAACACACTTGAGACCTTATACACA 840
QY 940 TATCTTTTGAAGTTTTCAGTCTCAACTGAAGCAGGTTATTTATGATAGTACGATGTC 999
DB 841 TATCTTTTGAAGTTTTCAGTCTCAACTGAAGCAGGTTATTTATGATAGTACGATGTC 900

QY 1000 AGAACACACAGATCAGTCTGCTGAAGGACCAACCAAACTGCTAACAGGCAACATCACA 1059
DB 901 AGAACACACAGATCAGTCTGCTGAAGGACCAACCAAACTGCTAACAGGCAACATCACA 960

QY 1060 GGAAAGTCTCTTTTCAATTTTATGGGACCCCAACCTATAGTAAACAGGGAATTTAGTTAT 1119
DB 961 GGAAAGTCTCTTTTCAATTTTATGGGACCCCAACCTATAGTAAACAGGGAATTTAGTTAT 1020

QY 1120 AGAGTTGAATTTATATGGACCATCAGGTCGCAATTTTGGATAACAGCACAAAAGACCTCAAG 1179
DB 1021 AGAGTTGAATTTATATGGACCATCAGGTCGCAATTTTGGATAACAGCACAAAAGACCTCAAG 1080

QY 1180 TTTGCAATTCATACTAACCTAACACATTTTACAATGTATGATGCTATATTTGGGCTGAAAC 1239
DB 1081 TTTGCAATTCATACTAACCTAACACATTTTACAATGTATGATGCTATATTTGGGCTGAAAC 1140

QY 1240 AGTCAGAGGACCTGGGCCCAAGTCAAAATTTTCAAGTATTCACCTCCACAGATGTCACAGG 1299
DB 1141 AGTCAGAGGACCTGGGCCCAAGTCAAAATTTTCAAGTATTCACCTCCACAGATGTCACAGG 1200

QY 1300 GCAGTGTGTTGATTTTCAACTTGCAGAGGTAGAATCCACGCAAGTAAGAATTAATCTTGGAA 1359
DB 1201 GCAGTGTGTTGATTTTCAACTTGCAGAGGTAGAATCCACGCAAGTAAGAATTAATCTTGGAA 1260

QY 1360 AAAACAAGCAACCAAAATGGAATTTTAAACCAATACCGAGTGAAGTGTCTAGTTCAGAG 1419
DB 1261 AAAACAAGCAACCAAAATGGAATTTTAAACCAATACCGAGTGAAGTGTCTAGTTCAGAG 1320

QY 1420 ACAGGAATTAATTTTGGAAATTAATTTGCTCACTGGAATAATAGATATATAATGACCC 1479
DB 1321 ACAGGAATTAATTTTGGAAATTAATTTGCTCACTGGAATAATAGAG---ATAATGACCC 1377

QY 1480 ATGGCTCCAGAAATTTGTGAACATAGTAGGCCAATGTTAGGATTTATAGAGGTTTCAGCA 1539
DB 1378 ATGGCTCCAGAAATTTGTGAACATAGTAGGCCAATGTTAGGATTTATAGAGGTTTCAGCA 1437

QY 1540 GAGATGTGCTGTGACCTTCACTCACTGTGTATATATTAACGCCATCAGATAA 1599
DB 1438 GAGATGTGCTGTGACCTTCACTCACTGTGTATATTAACGCCATCAGATAA 1497

QY 1600 AACTTTCTGCAAGGAATAGAGTGAAGACACAGCTTCAACAGTTGTAACTACAGGAAT 1659
DB 1498 AACTTTCTGCAAGGAATAGAGTGAAGACACAGCTTCAACAGTTGTAACTACAGGAAT 1557

QY 1660 CAGTATATTACTGACATTTGCAGCTGAAACAGCTGTCTTATGTTTATCAGGAGACTTGTACCT 1719
DB 1558 CAGTATATTACTGACATTTGCAGCTGAAACAGCTGTCTTATGTTTATCAGGAGACTTGTACCT 1617

QY 1720 TTTCACTGAGCAATGATTTAGTGTCTGCTTTTCACTCATGCTGGGAGAGGACCAACA 1779
DB 1618 TTTCACTGAGCAATGATTTAGTGTCTGCTTTTCACTCATGCTGGGAGAGGACCAACA 1677

QY 1780 GTTCTCAGTGTAGGACAGCTCAGCAAGTCCAGCTCCATTTAAATTTAACTATAA 1839
DB 1678 GTTCTCAGTGTAGGACAGCTCAGCAAGTCCAGCTCCATTTAAATTTAACTATAA 1737

QY 1840 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1899
DB 1738 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1797

QY 1900 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1959
DB 1798 AATATTAGTCTTCTCTATTTTATTTTGGGATCTTCCAGAAATATCCCAATGGAAA 1857

QY 1960 ACCATAGATAACAGCTTTCTCATAC-----AGGGTTAAAGAAATACACAAATACAAA 2013
DB 1958 ACCATAGATAACAGCTTTCTCATACAGGTATAGGGTTAAAGAAATACACAAATACAAA 1917
QY 2014 ATGAGAGTGGCAGCCTCAACCAGATGGAGAAAGTTCTTTGCTGGAAGAAATGACATC 2073
DB 1918 ATGAGAGTGGCAGCCTCAACCAGTGGAGAAAGTTCTTTGCTGGAAGAAATGACATC 1977
QY 2074 TTTGTGAGAACTTCAGAGAATGAACCGGAATCATCACCTCAAGATGTCGAAGTAATTGAT 2133
DB 1978 TTTGTGAGAACTTCAGAGAATGAACCGGAATCATCACCTCAAGATGTCGAAGTAATTGAT 2037
QY 2134 GTTACCCAGATGAATGAAGTTGAAGTGTCAACCAGGAAAGCCCAATGGGATCATTT 2193
DB 2038 GTTACCCAGATGAATGAAGTTGAAGTGTCAACCAGGAAAGCCCAATGGGATCATTT 2097
QY 2194 ATTGCTTATGAAGTGTATATAAAATATAGATACTTTATATATAGAGACATCAACA 2253
DB 2098 ATTGCTTATGAAGTGTATATAAAATATAGATACTTTATATATAGAGAAACATCAACA 2157
QY 2254 ACAGACATAATATTAAGGAATCTTAAGACCTCACACCTCTATAACATTTCTGTAAGTCT 2313
DB 2158 ACAGACATAATATTAAGGAATCTTAAGACCTCACACCTCTATAACATTTCTGTAAGTCT 2217
QY 2314 TACACAGATTTGCTCANGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG 2373
DB 2218 TACACAGATTTGCTCANGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG 2277
QY 2374 ACTGTGCTCATAGTGACACAGAAATATCACTTTACAAAATATTTCTTCGAGAGATT 2433
DB 2278 ACTGTGCTCATAGTGACACAGAAATATCACTTTACAAAATATTTCTTCGAGAGATT 2337
QY 2434 GAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAAATCAATAAAATATACAAATTTAT 2493
DB 2338 GAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAAATCAATAAAATATACAAATTTAT 2397
QY 2494 CTCAGAGAAATGAAGAAATGAGAAAGAACTATATAAATAACAACCTCTTTAAACCCAAAAC 2553
DB 2398 CTCAGAGAAATGAAGAAATGAGAAAGAACTATATAAATAACAACCTCTTTAAACCCAAAAC 2457
QY 2554 ATTAAAGTACTGAAGAAATATACCAATATATCATTTGAGGTGTCTGCTAGTACACTGAAA 2613
DB 2458 ATT-----CTGAAGAAATATACCAATATATCATTTGAGGTGTCTGCTAGTACACTGAAA 2511
QY 2614 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACGGAGGAAGATGCTCTGATTTCT 2673
DB 2512 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACGGAGGAAGATGCTCTGATTTCT 2571
QY 2674 CCCCTCAAGACTTCTCTGTAAGAAACAGTTGTCTGGTGTCAACGGTGAAGTTGTCTATGGCAA 2733
DB 2572 CCCCTCAAGACTTCTCTGTAAGAAACAGTTGTCTGGTGTCAACGGTGAAGTTGTCTATGGCAA 2631
QY 2734 CCACCTTGAGCCAAATGGAATATCTTTATATACAGTTTATGTCT---GGAATAGA 2790
DB 2632 CCACCTTGAGCCAAATGGAATATCTTTATATACAGTTTATGTCTTTATACAGTTTATGTCT 2691
QY 2791 TCATCATTAATAAATACTATTAATGTCACTGAAAACATCATTTGGAGTTATCAGATTTGGATTTAT 2850
DB 2692 TCATCATTAATAAATACTATTAATGTCACTGAAAACATCATTTGGAGTTATCAGATTTGGATTTAT 2751
QY 2851 AATGTTGAAATACAGTGTCTTATGTAACAGCTAGCACAGATTTGGTGTGATGGGAAAACAGGA 2910
DB 2752 AATGTTGAAATACAGTGTCTTATGTAACAGCTAGCACAGATTTGGTGTGATGGGAAAACAGGA 2811
QY 2911 AGCAATATCATTTAGCTTTCAACACAGAGGGAGCACAGGATCCTCCCAAGATGTT 2970
DB 2812 AGCAATATCATTTAGCTTTCAACACAGAGGG---ACCAAGCGATCCTCCCAAGATGTT 2868
QY 2971 TATTATGAAAACCTCAGTTCTTTCATCAATAATTTCTTTCTGGACACCTCCTTCAAACT 3030
DB 2869 TATTATGAAAACCTCAGTTCTTTCATCAATAATTTCTTTCTGGACACCTCCTTCAAACT 2928

QY 3031 AATGGGATTATACAAATATTACTCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCGAG 3090
DB 2929 AATGGGATTATACAAATATTACTCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCGAG 2988
QY 3091 AATTTTACACTCCATCAAACTAAACCAATGACTTTTGACAATATGACTCTATCCACAATTATA 3150
DB 2989 AATTTTACACTCCATCAAACTAAACCAATGACTTTTGACAATATGACTCTATCCACAATTATA 3048
QY 3151 GATAAACTGACAATATTACGTACTATATACATTTTGGTTAAACAGCAAGTACTTCAGTTGGA 3210
DB 3049 GATAAACTGACAATATTACGTACTATATACATTTTGGTTAAACAGCAAGTACTTCAGTTGGA 3108
QY 3211 AATGGGAATAAAGCAGTGAACATCATTTGAAGTATACACAGATCAAGACATACCTGAAGGG 3270
DB 3109 AATGGGAATAAAGCAGTGAACATCATTTGAAGTATACACAGATCAAGACGCTCCCTGAAGGG 3168
QY 3271 TTTGTTGGAACCTGACTTTACGAATCCATTTCTGTCAACTGCAATTAATGTAAGCTGGTC 3330
DB 3169 TTTGTTGGAACCTGACTTTACGAATCCATTTCTGTCAACTGCAATTAATGTAAGCTGGTC 3228
QY 3331 CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTTACAGCAGACT 3390
DB 3229 CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTTACAGCAGACT 3288
QY 3391 CCTCGCATGTGAGACCACTCTTGTTACATATGAGAGAAAGCATATATTTTGAATATCTG 3450
DB 3289 CCTCGCATGTGAGACCACTCTTGTTACATATGAGAGAAAGCATATATTTTGAATATCTG 3348
QY 3451 GAAAATACACATGATTATATATTAATAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3510
DB 3349 GAAAATACACATGATTATATATTAATAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3408
QY 3511 ACCTATCTGCCAGCTATACATCAAGACTGAAGAAGATGTCTCCAGAAACTTCACCAATA 3570
DB 3409 ACCTATCTGCCAGCTATACATCAAGACTGAAGAAGATGTCTCCAGAAACTTCACCAATA 3468
QY 3571 ATCAACACTTTTAAAAACCTTTCTCTACTACTCAGTTCTCTTATCATGGGATCCCCAGTA 3630
DB 3469 ATCAACACTTTTAAAAACCTTTCTCTACTACTCAGTTCTCTTATCATGGGATCCCCAGTA 3528
QY 3631 AAGCCAAATGGTCAATATATAGTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAA 3690
DB 3529 AAGCCAAATGGTCAATATATAGTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAA 3588
QY 3691 TCTTTTCACTTTCTTGATAATTTACATAATATTTGGAAGAGCTTTTACCACTTTTACATTTAT 3750
DB 3589 TCTTTTCACTTTCTTGATAATTTACATAATATTTGGAAGAGCTTTTACCACTTTTACATTTAT 3648
QY 3751 AGCTTTTGTGCTGCCAAGAACTAGAAAAGGACTTTGGTCTTCCAGATTTCTTTTCTTT 3810
DB 3649 AGCTTTTGTGCTGCCAAGAACTAGAAAAGGACTTTGGTCTTCCAGATTTCTTTTCTTT 3708
QY 3811 TACACAGATGAGTCAAGTCCGTTAGCACTCCACAAAATTTGACTTTTAACTGACTTACT 3870
DB 3709 TACACAGATGAGTCAAGTCCGTTAGCACTCCACAAAATTTGACTTTTAACTGACTTACT 3768
QY 3871 TCAGACTTTGTATGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3930
DB 3769 TCAGACTTTGTATGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA 3828
QY 3931 TATAGTTTAAATTCATGAACATGAACTGACATATATATATTAAGAATATATACAGGA 3990
DB 3829 TATAGTTTAAATTCATGAACATGAACTGACATATATATTAAGAATATATATACAGGA 3888
QY 3991 TTTTAAACCTGAAGCCAACTTTGTCGACTGGAAACAGTCAGCACCTACTCTATCCGTGTA 4050
DB 3889 TTTTAAACCTGAAGCCAACTTTGTCGACTGGAAACAGTCAGCACCTACTCTATCCGTGTA 3948
QY 4051 TCTGCTTCCACCAAGTTGGAAATGCGCAATCAATTTAGTTAAATGTAAGTAAATTTCAACACC 4110
DB 3949 TCTGCTTCCACCAAGTTGGAAATGCGCAATCAATTTAGTTAAATGTAAGTAAATTTCAACACC 4008
QY 4111 CAAGAATCAGTTTCCAGATGTCGTGCGAATATATGCAAGTGCATGGCAACTAGCTGGCAGTCA 4170

Qy	1033	CAAAAC	TCGCGTAA	CAGCGCA	ACATCA	CAGGAAA	AGTCCTTTT	CAATTTT	TATGGGAC	CCCA	CA	1091					
Db	946	CAAAAC	TCGCGTAA	CAGCGCA	CAATCA	CAGGAAA	AGTCCTTTT	CAATTTT	TATGGGAC	CCCA	CA	1005					
Qy	1093	ACTATAGT	TAAC	CAGGGA	AAATTTAG	TTATAG	AGTTGCA	ATTTATAT	TAGGAC	CACT--	-CAGT	CGC	1149				
Db	1006	ACTATAGT	TAAC	CAGGGA	AAATTTAG	TTATAG	AGTTGCA	ATTTATAT	TAGGAC	CACT	CAGCAG	GT	CGC	1065			
Qy	1150	ATTTTGGAT	TAA	CAGCACA	AAAAGAC	CTCA	AGTTTGC	ATTTTCA	CTAATCA	CTTAA	CACCA	TTTACA	1209				
Db	1066	ATTTTGGAT	TAA	CAGCACA	AAAAGAC	CTCA	AGTTTGC	ATTTTCA	CTAATCA	CTTAA	CACCA	TTTACA	1125				
Qy	1210	ATGTATGAT	GTCTAT	TATATG	CGGCTG	TAAACCA	GTGCGG	CAAGTGC	AGGAC	CTGGG	CCCAAGT	CTAAAT	TAT	1269			
Db	1126	ATGTATGAT	GTCTAT	TATATG	CGGCTG	TAAACCA	GTGCGG	CAAGTGC	AGGAC	CTGGG	CCCAAGT	CTAAAT	TAT	1185			
Qy	1270	TCAGTAT	TCACTCC	AC	CAGAGAT	GTTC	CAGGGG	CAGTGTTG	ATTTTACA	ATTTTACA	ATTTTACA	ATTTTACA	1329				
Db	1186	TCAGTAT	TCACTCC	AC	CAGAGAT	GTTC	CAGGGG	CAGTGTTG	ATTTTACA	ATTTTACA	ATTTTACA	ATTTTACA	1245				
Qy	1330	GAATCC	CAGCA	GAATTA	CTACTTTG	GAAGAAA	CCAGCACA	CAACCA	ATGGA	ATTTAT	TTAA	CT	1389				
Db	1246	GAATCC	CAGCA	GAATTA	CTACTTTG	GAAGAAA	CCAGCACA	CAACCA	ATGGA	ATTTAT	TTAA	CT	1305				
Qy	1390	CAATAC	CGAGT	GAAGT	GTCTAG	TTCC	CAGAGAC	AGGAATA	TTTTTGG	AAAA	TACTTTG	CTC	1449				
Db	1306	CAATAC	CGAGT	GAAGT	GTCTAG	TTCC	CAGAGAC	AGGAATA	TTTTTGG	AAAA	TACTTTG	CTC	1365				
Qy	1450	ACTGGAA	ATATG	AGTATAT	TAATATG	ACCCAT	GGCTCC	AGAAAT	TGTGA	ACATAG	TAGAG	1509					
Db	1366	ACTGGAA	ATATG	AGTATAT	TAATATG	ACCCAT	GGCTCC	AGAAAT	TGTGA	ACATAG	TAGAG	1422					
Qy	1510	CCAATGG	TAGGAT	TATATG	AGGGT	TTCA	GCAGCAG	AGATGTCTG	CTGAC	CTTCA	CTCA	TTGCT	1569				
Db	1423	CCAATGG	TAGGAT	TATATG	AGGGT	TTCA	GCAGCAG	AGATGTCTG	CTGAC	CTTCA	CTCA	TTGCT	1482				
Qy	1570	ACATTT	TATATATA	ACGCCAT	CCAGAT	TA	AAAACTTTCTG	CAAGGA	ATAGAG	CTGGAAG	AC	1629					
Db	1483	ACATTT	TATATATA	ACGCCAT	CCAGAT	TA	AAAACTTTCTG	CAAGGA	ATAGAG	CTGGAAG	AC	1542					
Qy	1630	CAGACT	TTCA	CCAGT	TTGTA	CTACA	AGGAATC	AGTATAT	TACTG	ACAT	TTGCAG	CTGAC	1689				
Db	1543	CAGACT	TTCA	CCAGT	TTGTA	CTACA	AGGAATC	AGTATAT	TACTG	ACAT	TTGCAG	CTGAC	1602				
Qy	1690	CTGTCT	TTATGTTAT	CAGCAG	ACTTGT	TACCTTTCA	CTGAGCACA	TGATTTAG	TGTAT	CTTGCT	1749						
Db	1603	CTGACT	TTATGTTCTT	ATCAG	ATTAAG	AGGATTTTGG	GCTGAGAC	CAATTTGG	GGTTTTCT	TAGA	1662						
Qy	1750	TTCA	CCATCAT	TGGGAG	AGGACCA	CCAC	AGTCTCTC	AGTGT	TAGGAC	ACGT	CAGCAAG	GTG	1809				
Db	1663	TATACA	ATCATG-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1710				
Qy	1810	CCAAGCT	CCATTA	AAAA	TTATAA	ACTATA	AAAA	TTATAG	TTCTTCA	CTCTAT	TTTGT	TATAT	1869				
Db	1711	CTTTTGT	CAGCCCCAAA	TTT	CAGAG	TTTACA	CATGTTTAC	CAATTA	CCAGAA	GAATTTT	TTTAC	AC	1770				
Qy	1870	TGGCAT	CTCC	CAGATA	TCCCA	ATGGA	AAAA	TTAATCTCA	TATAC	GTATAT	TATG	CAATG	GA	1929			
Db	1771	TGGGAT	CCTC	CAGAT	CTCTG	TATTTTTT	TCA	TCATTA	CTTAT	TCATAT	TTTGG	ATG	TT	1830			
Qy	1930	TTGATAT	CAAA	CAGAG	CAATCC	CAGATA	AACTAC	CA	TAGATA	TAA	CAGCTTTCT	CA	TAA	CA	1989		
Db	1831	AA	CAATCC	AGAGT	TATTTTTT	TAAG	GACATTTAA	CAGT	TTTGT	CTCTCT	TTGCTTAT	TATAGG	1890				
Qy	1990	TTAA	AGAA	ATACA	CAAA	ATACA	AAATG	AGAG	GTGG	CGCCTCA	ACCC	ACGAT	TGG	GA	2049		
Db	1891	TTAA	AGAA	ATACA	CAAA	ATACA	AAATG	AGAG	GTGG	CGCCTCA	ACCC	ACGAT	TGG	GA	1950		
Qy	2050	TC	TTTGT	CTGA	AGAA	ATG	CATCTTTG	TG	GA	ACTTT	CAGA	AGAT	GA	CCG	AT	CA	2109
Db	1951	TC	TTTGT	CTGA	AGAA	ATG	C										

2011	CCTCAAGATGCGAAGTAATTGATTTACGCAGATAAAATAAGTGTGAAGTGTCACCA	2070
2170	CCGGAAAAGCCCAATGGGATCATTTATGCTTTATGAAGTCTATATAAAAAATATAGACT	2229
2071	CCGGAAAAGCCCAATGGGATCATTTATGCTTTATGAAGTCTATATAAAAAATAGACT	2130
2230	TTATATATGAAGAACATCAACAACAGACATATATATTAAGGAACTTTAAGACCCTCAC	2289
2131	TTATATATGAAGAACATCAACAACAGACATATATTAAGGAACTTTAAGACCCTCAC	2190
2290	CTCTATACATTTCTGTATAGGTCCTTACACAGATTTGGTCA TGGCAATCAGGTATCTTCT	2349
2191	CTCTATAACATTTCTGTATAGGTCCTTACACAGATTTGGTCA TGGCAATCAGGTATCTTCT	2250
2350	TTACTCTCTGTAAAGGACTTCGGAGACTGTGCTGATGTGCA CAGAAAAATATCACTTAC	2409
2251	TTACTCTCTGTAAAGGACTTCGGAGCTCAGTGCCTGATAGT GCACAGAAAAATATCNC TTAC	2310
2410	AAAAATATTTCTTCGGAGAGATTGAGCTATCATCTCTTCCC CAAGTAGTCCCAATGGGA	2469
2311	AAAAATATTTCTTCGGAGAGATTGAGCTATCATCTCTTCCC CAAGTAGTCCCAATGGGA	2370
2470	ATCATAAAAAATATACAAATTTATCTCAAGAGAGATTAAT GGAAATGAGGAAGA CTATA	2529
2371	ATCATAAAAAATATACAAATTTATCTCAAGAGAGATTAAT GGAAATGAGGAAGA CTATA	2430
2530	AATACAACTCTTTAAACCCAAAACATTTAAAGTACTGGA GAAATATACC CAATATATCATT	2589
2431	AATACAACTCTTTAAACCCAAAACATTTAAAGGTCGNA GAATATACC CAATATATCATT	2490
2590	GAGGTGCTCTGCTAGTACATGCAAGAGTGGAAGGATT CGGAGTCTGCCAT AAGTATATCTG	2649
2491	GAGGTGCTCTGCTAGTACATCAAAGGTGAAGGATTCGG AGTCTGCCAT AAGTATATCTG	2550
2650	ACGGAGGAAGATGCTCTGATTTCTCCCCCTCAAGACTTCT GTGTA AAAACAGTGTCTCTGGT	2709
2551	ACGGAGGAAGATGCTCTGATTTCTCCCCCTCAAGACTTCT GTGTA AAAACAGTGTCTCTGGT	2610
2710	GTCAAGGTGAAGTTGT CATGSCAACACACCCCTGGAGC CAAATGGAAATTTATCCTTTATTTAC	2769
2611	GTCAAGGTGAAGTTGT CATGSCAACACACCCCTGGAGC CAAATGGAAATTTATCCTTTATTTAC	2670
2770	ACAGTTTATGTCTGGAA TAGATCATATTA AAAA ACTPATTAATGTCACTG AAACATCATTTG	2829
2671	ACAGTTTATGTCTGG --- AGATCATATTA AAAA ACTPATTAATGTCACTG AAACATCATTTG	2727
2830	GAGTTATCAGATTTGGATTTAATGTTGAATACAGTGCCTT ATGTTAA CAGCTAGCACCAGA	2889
2728	GAGTTATCAGATTTGGATTTAATGTTGAATACAGTGCCTT ATGTTAA CAGCTAGCACCAGA	2787
2890	TTTTGGTGAATGGAAAA CAGGAAGCAATATCATTTAGCTT TC AAAACAC CAGAGGGAGCACA	2949
2788	TTTTGGTGAATGGAAAA CAGGAAGCAATATCATTTAGCTT TC AAAACAC CAGAGGG --- ACCA	2844
2950	AGCGATCTCTCCCAAGATGTTTATTTATGCAAA CCTCAGT TCTTCATCAATTAATTTCTTTTC	3009
2845	AGCGATCTCTCCCAAGATGTTTATTTATGCAAA CCTCAGT TCTTCATCAATTAATTTCTTTTC	2904
3010	TGGACACTCTCTTCAA AACCTTAATGGGA TTATACAAATATTA CTCTGTTTATTTACAGAAAT	3069
2905	TGGACACTCTCTTCAA AACCTTAATGGGA TTATACAAATATTA CTCTGTTTATTTACAGAAAT	2964
3070	ACTTCAGGTATTTTTATG SCAGAA TTTTATACATCTCCA TGAACCAACCAATGCTTTTGACAAT	3129
2965	ACTTCAGGTATTTTTATG SCAGAA TTTTATACATCTCCA TGAACCAACCAATGCTTTTGACAAT	3024
3130	ATGACTGTATCCCAAA TTTATAGATAAA ACTGACAAATATTCAGCTACTATACATTTTGGTTA	3189
3025	ATGACTGTATCCCAAA TTTATAGATAAA ACTGACAAATATTCAGCTACTATACATTTTGGTTA	3084
3190	ACAGCAAGTACTTCA GTTTGGAAATGGGAATAAAGCAGTGACATCATTTGAA GTATACACA	3249

DB	3085	ACAGCAGTACTTCAGTTGGAAATGGGGATAAAAGCAGTGACATCAATTGAAGTATATACACA	3144
QY	3250	GATCAAGACATACCTGAAGGGTTTGTGGAAACCTGACTTACGAAATCCATTTTCGTCAACT	3309
DB	3145	GATCAAGAGGTACCTGAAGGGTTGTGGAAACCTGACTTACGAAATCCATTTTCGTCAACT	3204
QY	3310	GCAATAATGTAACTGGTCCCAACGGCTCAACAAACGGTCTAGTCTTCTACTATGTT	3369
DB	3205	GCAATAATGTAAAGCTGGTCCCAACGGCTCAACAAACGGTCTAGTCTTCTACTATGTT	3264
QY	3370	TCAGTGAATCTTACAGCAGACTCCTCGCCATGTGAGACACCTCTGTGTTTACATATGACAGA	3429
DB	3265	TCAGTGAATCTTACAGCAGACTCCTCGCCATGTGAGACACCTCTGTGTTTACATATGACAGA	3324
QY	3430	AGCATATATTTTGATAATCTGGAAAAATACACTGATTATATATTATATAAAATTTACTCCATCA	3489
DB	3325	AGCATATATTTTGATAATCTGGAAAAATACACTGATTATATATTATATAAAATTTACTCCATCA	3384
QY	3490	ACAGAAAAGGATTTCTCTGATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAAGAT	3549
DB	3385	ACAGAAAAGGATTTCTCTGATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAAGAT	3444
QY	3550	GTCCACAGAACTTCACCAATAATAACACACTTTTAAABACCTTTCTCTACTCCTCAGTCTCTC	3609
DB	3445	ATCCAGAAACTTCACCAATAATAACACACTTTTAAABACCTTTCTCTACTCCTCAGTCTCTC	3504
QY	3610	TTATCATGGGATCCCCAGTAAAGCCAAATGGTCCAATAATAAGTTATGATTTAAACTTTTA	3669
DB	3505	TTATCATGGGATCCCCAGTAAAGCCAAATGGTCCAATAATAAGTTATGATTTAAACTTTTA	3564
QY	3670	CAAGGACAAATGAAAAATTATCTTTCAATTACTTCTGATTAATATACATAATTTGGAAGAG	3729
DB	3565	CAAGGACAAATGAAAAATTATCTTTCAATTACTTCTGATTAATATACATAATTTGGAAGAG	3624
QY	3730	CTTTTCAACATTTACATTTATAGCTTTTTCGTCGCGCAAGACTGAAAAGGACATTTGGT	3789
DB	3625	CTTTTCAACATTTACATTTATAGCTTTTTCGTCGCGCAAGACTGAAAAGGACATTTGGT	3684
QY	3790	CCTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCAGTGCCTTTAGCACCTCCACAATAAT	3849
DB	3685	CCTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCAGTGCCTTTAGCACCTCCACAATAAT	3744
QY	3850	TTGACTTTAATCAACTGTACTTTCAGACTTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTT	3909
DB	3745	TTGACTTTAATCAACTGTACTTTCAGACTTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTT	3804
QY	3910	CCAGTGGTATTTGTTAAAGTATATAGTTTAAATTTCAATGAACTGAACTGACACTATA	3969
DB	3805	CCAGTGGTATTTGTTAAAGTATATAGTTTAAATTTCAATGAACTGAACTGACACTATA	3864
QY	3970	TATTATAAGAAATATATACGATTTTAAAACTGAAAGCCAAACTCTGTGGACTGGAAACGAGTC	4029
DB	3865	TATTATAAGAAATATATACGATTTTAAACTGAAAGCCAAACTCTGTGGACTGGAAACGAGTC	3924
QY	4030	AGCACCTACTCTATCCGTGTATCTGCGTTCAACAAAGTTGGAAATGGCAATCAATTTAGT	4089
DB	3925	AGCACCTACTCTATCCGTGTATCTGCGTTCAACAAAGTTGGAAATGGCAATCAATTTAGT	3984
QY	4090	AATGTAGTAAATTCACAAACCAAGATCAGTTCAGATGTCGTGCGAGATATGCGAGTGC	4149
DB	3985	AATGTAGTAAATTCACAAACCAAGATCAGTTCAGATGTCGTGCGAGATATGCGAGTGC	4044
QY	4150	ATGGCAACTAGCTGGCAGTCAGTTTATGCGAAATGGGATCCACCCAAAAGGCCAAATGGA	4209
DB	4045	ATGGCAACTAGCTGGCAGTCAGTTTATGCGAAATGGGATCCACCCAAAAGGCCAAATGGA	4104
QY	4210	ATAATAACGCAGTATATGGTAAACAGTTGAAAGGAAATTTCTACAAAAGTTTCTCCCCCAAGAT	4269
DB	4105	ATAATAACGCAGTATATGGTAAACAGTTGAAAGGAAATTTCTACAAAAGTTTCTCCCCCAAGAT	4164
QY	4270	CACATGTACATTTTCATAAAAGCTTTTGGCCAACTCTCATATGTCTTTTAAAGTAAGAGCT	4329
DB	4165	CACATGTACATTTTCATAAAAGCTTTTGGCCAACTCTCATATGTCTTTTAAAGTAAGAGCT	4224

Db 4 TTAGGAGGAACTAACAAACAGAAAGGACATCCACCAAAACCCTCTGTATCATCAC 63
QY 1189 ACTAACCTAACCACTTTACAATGTATGATGCTATATTGCGGCTGAAACCAAGTGCAGGG 1248
Db 64 CATCATCAAGACCAAAAGTAGATAAAACCAAAAGATGGGAAAAACAGAGCAAAAA 123
QY 1249 ACTGGCCCCAAGTCAATATTTCAGTATTTCATCTCCACAGATGTTCCAGGGCAGTGT 1308
Db 124 ACTGGAAACTCTAAAGCAGAGCACCTCTCTCTCCCAAGGATGGCAGGGCAGTGT 183
QY 1309 GATTTACAACTTGCAGAGGTAGAAATCCAGCAAGTAAAGAAATACCTTGGAGAAACCAACGA 1368
Db 184 GATTTACAACTTGCAGAGGTAGAAATCCAGCAAGTAAAGAAATACCTTGGAGAAACCAACGA 243
QY 1369 CAAACCAATGGAATTAATTAACCAATACCGAGTGAAGTGTCTAGTTCACAGACAGGAATA 1428
Db 244 CAAACCAATGGAATTAATTAACCAATACCGAGTGAAGTGTCTAGTTCACAGACAGGAATA 303
QY 1429 ATTTTGGAAATACCTTGTCTCACTGCAATATGAGTATATAAATGACCCCATGGCTCCA 1488
Db 304 ATTTTGGAAATACCTTGTCTCACTGCAATATGAGTATATAAATGACCCCATGGCTCCA 363
QY 1489 GAAATGTCAATAGTAGGCAATGTAGGATATATGAGGTTTACGACAGAGATGTCG 1548
Db 364 GAAATGTCAATAGTAGGCAATGTAGGATATATGAGGTTTACGACAGAGATGTCG 423
QY 1549 TCTGACCTTCACTCACTTCTACATTTATATATACAGCCATCCAGATAAAATCTTCT 1608
Db 424 TCTGACCTTCACTCACTTCTACATTTATATATACAGCCATCCAGATAAAATCTTCT 483
QY 1609 GCAAGGATAGAGCTGAACACGACTTCCACGTTGTAACTACAGGAATCAGTATATT 1668
Db 484 GCAAGGATAGAGCTGAACACGACTTCCACGTTGTAACTACAGGAATCAGTATATT 543
QY 1669 ACTGACATTCAGCTGAACAGCTGCTTATGTTATTCAGGAGACTTGTACTTTCACTGAG 1728
Db 544 ACTGACATTCAGCTGAACAGCTGCTTATGTTATTCAGGAGACTTGTACTTTCACTGAG 603
QY 1729 CACATGATTAGTGTCTGCTTTTCACTCATGGGAGAGGACCAACACAGTTCTCAGT 1788
Db 604 CACATGATTAGTGTCTGCTTTTCACTCATGGGAGAGGACCAACACAGTTCTCAGT 663
QY 1789 GTTAGGACAGCTCAGCAAGTGCAGCTCCATTTAAATTTATAAATTAATAAATATTAGT 1848
Db 664 GTTAGGACAGCTCAGCAAGTGCAGCTCCATTTAAATTTATAAATTTATAAATATTAGT 723
QY 1849 TCTTCATCTATTGTTTATATTGGGATCTCCAGATATCCCAATGGGAAATTAATCTAC 1908
Db 724 TCTTCATCTATTGTTTATATTGGGATCTCCAGATATCCCAATGGGAAATTAATCTAC 783
QY 1909 TATACGATTTATGCAATGGAAATGGATACAAACAGAGCAATCCAGATATACCATAGAT 1968
Db 784 TATACGATTTATGCAATGGAAATGGATACAAACAGAGCAATCCAGATATACCATAGAT 843
QY 1969 AACAGCTTTCTCATACAGGTTAAAGAAATACAAAAATACAAATGAGAGTGGCAGCC 2028
Db 844 AACAGCTTTCTCATACAGGTTAAAGAAATACAAAAATACAAATGAGAGTGGCAGCC 903
QY 2029 TCAACCCAGATGGAGAAATGTTCTGCTGGAAGAAATGACATCTTGTGGAACCTTCA 2088
Db 904 TCAACCCAGATGGAGAAATGTTCTGCTGGAAGAAATGACATCTTGTGGAACCTTCA 963
QY 2089 GAAGATGAACCGGAATCATCACTCAAGATGTGGAAGTAAATGATGTTTACCCGAGATGAA 2148
Db 964 GAAGATGAACCGGAATCATCACTCAAGATGTGGAAGTAAATGATGTTTACCCGAGATGAA 1023
QY 2149 ATAAGTTGAAGTGTACCACCCGAAAGCCCAATGGGATCATTTATGCTTTATGAAGTG 2208
Db 1024 ATAAGTTGAAGTGTACCACCCGAAAGCCCAATGGGATCATTTATGCTTTATGAAGTG 1083
QY 2209 CTATATAAAATATAGATCTTTTATATATGAAGACACATCAACACAGACATATATTA 2268
Db 1084 CTATATAAAATATAGATCTTTTATATATGAAGACACATCAACACAGACATATATTA 1143

QY 2269 AGGAACCTTAAGACCTTCACACCCCTCTATATAACATTTCTGTAAAGGTCTTTACACCAAGTTGGT 2328
Db 1144 AGGAACCTTAAGACCTTCACACCCCTCTATATAACATTTCTGTAAAGGTCTTTACACCAAGTTGGT 1203
QY 2329 CATGGCAATCAGGTATCTTCTTACTCTCTGTAAAGACTTCGGAGACTTCGGCTGATAGT 2388
Db 1204 CATGGCAATCAGGTATCTTCTTACTCTCTGTAAAGACTTCGGAGACTTCGGCTGATAGT 1263
QY 2389 GCACAGAAAATATCACTTTACAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCCCTT 2448
Db 1264 GCACAGAAAATATCACTTTACAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCCCTT 1323
QY 2449 CCCCAGATGATGCCAATGGAATCATAAAAAATATACAAATTTATCTCAAGAGAAGTAAT 2508
Db 1324 CCCCAGATGATGCCAATGGAATCATAAAAAATATACAAATTTATCTCAAGAGAAGTAAT 1383
QY 2509 GGAATAGGAAAGAACTATATAACCTCTTTTAAACCCAAAACATTAAGTACTGGAAG 2568
Db 1384 GGAATAGGAAAGAACTATATAACCTCTTTTAAACCCAAAACATTAAGTACTGGAAG 1443
QY 2569 AAATATACCAATATATCACTGAGGTGCTGCTAGTACACTGAAAGGTGAAGGATTCGG 2628
Db 1444 AAATATACCAATATATCACTGAGGTGCTGCTAGTACACTGAAAGGTGAAGGATTCGG 1503
QY 2629 AGTGTCTCCATAAGTATATCTGACGGAAGATGCTCTGTGATTTCTCCCTCAAGACTTC 2688
Db 1504 AGTGTCTCCATAAGTATATCTGACGGAAGATGCTCTGTGATTTCTCCCTCAAGACTTC 1563
QY 2689 TCTGTAAAACAGTTGCTGTGTGCTGACGGAAGTGTGATGCGAACCAACCCCTGGAGCCA 2748
Db 1564 TCTGTAAAACAGTTGCTGTGTGCTGACGGAAGTGTGATGCGAACCAACCCCTGGAGCCA 1623
QY 2749 AATGGAATTTATCTTTTATACACAGTTTATGTCTGGAATAGATCATTTAAACACTATT 2808
Db 1624 AATGGAATTTATCTTTTATACACAGTTTATGTCTGGAATAGATCATTTAAACACTATT 1583
QY 2809 AATGTCACTGAAACATCATTTGGAGTTATCAGATTTTGGATTTAATGTTGAATACAGTGT 2868
Db 1684 AATGTCACTGAAACATCATTTGGAGTTATCAGATTTTGGATTTAATGTTGAATACAGTGT 1743
QY 2869 TATGTAAACAGCTAGCACCAAGATTTGGTGTGGAAGAAACAGGAAGCAATATCATTTAGCTTT 2928
Db 1744 TATGTAAACAGCTAGCACCAAGATTTGGTGTGGAAGAAACAGGAAGCAATATCATTTAGCTTT 1803
QY 2929 CAAACACAGAGGGAGCACCAAGCATCTCTCCAAAGATGTTTATTTATCAAACTCAGT 2988
Db 1804 CAAACACAGAGGGAGCACCAAGCATCTCTCCAAAGATGTTTATTTATGCAAACCTCAGT 1863
QY 2989 TCTTCATCAATAATTTCTTTCTGGACACCTCTCTCAAAACCTTAATGGGATTTACAAATAT 3048
Db 1864 TCTTCATCAATAATTTCTTTCTGGACACCTCTCTCAAAACCTTAATGGGATTTACAAATAT 1923
QY 3049 TACTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGCAAGATTTTACATCCATGAA 3108
Db 1924 TACTCTGTTTATTAACAGAAATACCTTCAGGTACTTTTATGCAAGATTTTACATCCATGAA 1983
QY 3109 CTAACCAATGCTTTGACAAATATGACTGTATCCAAATTTATAGATAAACTGACAAATTTTC 3168
Db 1984 GTAACCAATGCTTTGACAAATATGACTGTATCCAAATTTATAGATAAACTGACAAATTTTC 2043
QY 3169 AGCTACTATACATTTTGGTTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAATAAAAGCAGT 3228
Db 2044 AGCTACTATACATTTTGGTTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAATAAAAGCAGT 2103
QY 3229 GACATCATGAGATATACAGATCAAGACATACCTGGAAGGTTTGTGTTGGAACCTGACT 3288
Db 2104 GACATCATGAGATATACAGATCAAGACATACCTGGAAGGTTTGTGTTGGAACCTGACT 2163
QY 3289 TACGAATCCATTTCTGCACTGCAATTAATGTAAGCTGGTCCCAACCGCTCAACCAAC 3348
Db 2164 TACGAATCCATTTCTGCACTGCAATTAATGTAAGCTGGTCCCAACCGCTCAACCAAC 2223

QY	3349	GGTCTAGTCTCTTCTACTATATGTTTCACTGATCTTACAGCAGACTCTCGCCACTGTGAGACCA	3408		Db	3304	CAGTCACTAGTGCACCACTTGACATGGATAGACCTGACACTATCTCTGGCTACTTTCAA	3363	
Db	2224	GGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACTCTCGCCATGTGAGACCA	2283		QY	4489	AATTTACAAATTTACCACTTCTCGTCTCAAAAATGCAAAAGATGGGAATCCGGAAGAA	4548	
QY	3409	CCTCTTGTGTACATATGAGAGAGCATATATTTTGTGAATCTGGAATAATACACTGATTTAT	3468		Db	3364	AATTTACAAATTTACCACTTCTCGTCTCAAAAATGCAAAAGATGGGAATCCGGAAGAA	3423	
Db	2284	CCTCTTGTGTACATATGAGAGAGCATATATTTTGTGAATCTGGAATAATACACTGATTTAT	2343		QY	4549	TCGTGTGATATCAAAAATTTCAATACCTCTATGAGCTCACTTAACCTGAAGACAGTA	4608	
QY	3469	ATATTTAAATTTACTCTCCATCAACAGAAAAGGATTTCTCTGATACCTATATCTGCCAGCTA	3528		Db	3424	TCGTGTGATATCAAAAATTTCAATACCTCTATGAGCTCACTTAACCTGAAGACAGTA	3483	
Db	2344	ATATTTAAATTTACTCTCCATCAACAGAAAAGGATTTCTCTGATACCTATATCTGCCAGCTA	2403		QY	4609	TATGGATTTAAAGAAATTTAGATGGTATAGATTTCCAAGTGGCTGCCAGCACCAATGTGGC	4668	
QY	3529	TACATCAAGACTGAAGAGATGTCACAGAACTTCCACCAATATCAACACTTTTAAARAC	3588		Db	3484	TATGGATTTAAAGAAATTTAGATGGTATAGATTTCCAAGTGGCTGCCAGCACCAATGTGGC	3543	
Db	2404	TACATCAAGACTGAAGAGATGTCACAGAACTTCCACCAATATCAACACTTTTAAARAC	2463		QY	4669	TATGGCAATGCTTTCAAACTGGATTTCTACAAAACTCTGCCCTGGCCCTCCAGATGGTCT	4728	
QY	3589	CTTTCTCTACCTCAGTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGAATA	3648		Db	3544	TATGGCAATGCTTTCAAACTGGATTTCTACAAAACTCTGCCCTGGCCCTCCAGATGGTCT	3603	
Db	2464	CTTTCTCTACCTCAGTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGAATA	2523		QY	4729	CCTGAAATGTTTCATGTAGTAGCAACATCACCTTTTAGCATCAGCATAGCTGGAGTGA	4788	
QY	3649	ATAAGTTATGATTTAACTTTACAGGACCAATATGAATAATTTCTTCAATTAATCTTGAT	3708		Db	3604	CCTGAAATGTTTCATGTAGTAGCAACATCACCTTTTAGCATCAGCATAGCTGGAGTGA	3663	
Db	2524	ATAAGTTATGATTTAACTTTACAGGACCAATATGAATAATTTCTTCAATTAATCTTGAT	2583		QY	4789	CCTGCTGTCTATTTACTGGACCAACATGTTTATCTGATTTGATGTCAAAATCGGTAGATATGAT	4848	
QY	3709	AATTACATATATTGGAAGAGCTTTCCACCAATTTACATATATATAGCTTTTTCGTCGCGCA	3768		Db	3664	CCTGCTGTCTATTTACTGGACCAACATGTTTATCTGATTTGATGTCAAAATCGGTAGATATGAT	3723	
Db	2584	AATTACATATATTGGAAGAGCTTTCCACCAATTTACATATATATAGCTTTTTCGTCGCGCA	2643		QY	4849	GAATTTAATATATCTTTCATCAAGTCAAAATGAAGAAAAATAAAACATAGAAATTAAGAT	4908	
QY	3769	AGAACTAGAAAAAGCACTTGGTCTCTCCAGTATTTCTTTTCTTTTACACAGATGAGTCAGTG	3828		Db	3724	GAATTTAATATATCTTTCATCAAGTCAAAATGAAGAAAAATAAAACATAGAAATTAAGAT	3783	
Db	2644	AGAACTAGAAAAAGCACTTGGTCTCTCCAGTATTTCTTTTCTTTTACACAGATGAGTCAGTG	2703		QY	4909	TTAGAAATTTTCAAGGATTTCTGTAGTGATCACTGCAATTTACTGGGAACATTAAGTCT	4968	
QY	3829	CGTTAGCACTTCCACAAATTTGACTTTTAACTCAACTGACTGACTTTCAGACTTTGATGGCTG	3888		Db	3784	TTAGAAATTTTCAAGGATTTCTGTAGTGATCACTGCAATTTACTGGGAACATTAAGTCT	3843	
Db	2704	CGTTAGCACTTCCACAAATTTGACTTTTAACTCAACTGACTTTCAGACTTTGATGGCTG	2763		QY	4969	GCATATGTAGAAAGGAAGTCAAGTCTGAAATGATTTGTTTACTTACTTTAGAAATCAGCCCCA	5028	
QY	3889	AAATGAGCCCAAGTCTCTCCAGTGGTATTTAAAGTATATAGTTTAAATTTCAAT	3948		Db	3844	GCATATGTAGAAAGGAAGTCAAGTCTGAAATGATTTGTTTACTTACTTTAGAAATCAGCCCCA	3903	
Db	2764	AAATGAGCCCAAGTCTCTCTCCAGTGGTATTTAAAGTATATAGTTTAAATTTCAAT	2823		QY	5029	AAGGACCCACCTTAAACAAATGACATTTTCAAGAGATACAGATGAAGTTTCAAAAATTTCAA	5088	
QY	3949	GAACATGAACCTGACACTATATATTAAGAAATATATCAGGATTTAAATCTGAAGCCAAA	4008		Db	3904	AAGGACCCACCTTAAACAAATGACATTTTCAAGAGATACAGATGAAGTTTCAAAAATTTCAA	3963	
Db	2824	GAACATGAACCTGACACTATATATTAAGAAATATATCAGGATTTAAATCTGAAGCCAAA	2883		QY	5089	TTAAGTCTCTTCTCTCTCTCAACCTATGGAATATCCAAAGTATATCAAGTATATCAAGCTCTGTT	5148	
QY	4009	CTTGTGGACTGGAACCACTGACGACCTACTCTATCCGTGTATCTCGGTTTACCAAGTT	4068		Db	3964	TTAAGTCTCTTCTCTCTCTCAACCTATGGAATATCCAAAGTATATCAAGTATATCAAGCTCTGTT	4023	
Db	2884	CTTGTGGACTGGAACCACTGACGACCTACTCTATCCGTGTATCTCGGTTTACCAAGTT	2943		QY	5149	TACCGAAGAGATGATCTCTACTCTGTCCAGATTTCAACCTCAGTATTTATACAGAAACC	5208	
QY	4069	GGAAATGGCAATCAATTTTATGTAATGTAGTAAATTTCAACCCCAAGAAATCAGTTCCAGAT	4128		Db	4024	TACCGAAGAGATGATCTCTACTCTGTCCAGATTTCAACCTCAGTATTTATACAGAAACC	4083	
Db	2944	GGAAATGGCAATCAATTTTATGTAATGTAGTAAATTTCAACCCCAAGAAATCAGTTCCAGAT	3003		QY	5209	AACACATTCGTCTTTCGAATGTAGAAAGCACTTAAAGGTGGACATACATACATATCAGT	5268	
QY	4129	GTCTGTCAGAAATATGAGTGCATGGCAACTAGCTGGCAGTCACTTTTATGTAATTTGGAT	4188		Db	4084	AACACATTCGTCTTTCGAATGTAGAAAGCACTTAAAGGTGGACATACATACATATCAGT	4143	
Db	3004	GTCTGTCAGAAATATGAGTGCATGGCAACTAGCTGGCAGTCACTTTTATGTAATTTGGAT	3063		QY	5269	GTTTACGCACTCAATAGTCTGTGTCAGGTCCTCAAGGTTCCGATGAGATTAACCATGGAT	5328	
QY	4189	CCACCCAAAGGCAATGGAATAATAACGCAATATATGGTAAACAGTTGAAGGAATTTCT	4248		Db	4144	GTTTACGCACTCAATAGTCTGTGTCAGGTCCTCAAGGTTCCGATGAGATTAACCATGGAT	4203	
Db	3064	CCACCCAAAGGCAATGGAATAATAACGCAATATATGGTAAACAGTTGAAGGAATTTCT	3123		QY	5329	ATCAAGCTTCCAGCAGCAACAAAAACCAACCCCAACCCCTATTTTATGATGCCAGGAAAA	5388	
QY	4249	ACAAAGTTTCTCCCAAGATACATGACATTTTCAATAAGCTTCTTGCATACCTCA	4308		Db	4204	ATCAAGCTTCCAGCAGCAGCAACAAAAACCAACCCCAACCCCTATTTATGATGCCAGGAAAA	4263	
Db	3124	ACAAAGTTTCTCCCAAGATACATGACATTTTCAATAAGCTTCTTGCATACCTCA	3183		QY	5389	CTGCTTGTGACTTCAACAAATTTACAAATCAGAAATGCCAAATGTTTACTACAGTATGAT	5448	
QY	4309	TATGTCTTTTAAAGTAAAGCTTCAACCTCAGCTGGTGAAGGTTGATGAAGACCATGCCAT	4368		Db	4264	CTGCTTGTGACTTCAACAAATTTACAAATCAGAAATGCCAAATGTTTACTACAGTATGAT	4323	
Db	3184	TATGTCTTTTAAAGTAAAGCTTCAACCTCAGCTGGTGAAGGTTGATGAAGACCATGCCAT	3243		QY	5449	CATGACCAATTAATAATGTACAAGTCTTTCGACAGAAACAGGAGCTCAGCATGATGGA	5508	
QY	4369	GTGAGCACTACTCTGAAACAGTTTCCAGTGTTCACAAATATTTGCTTTTCTGATGTT	4428		Db	4324	CATGACCAATTAATAATGTACAAGTCTTTCGACAGAAACAGGAGCTCAGCATGATGGA	4383	
Db	3244	GTGAGCACTACTCTGAAACAGTTTCCAGTGTTCACAAATATTTGCTTTTCTGATGTT	3303		QY	5509	AATGTAAACAAAGTGTGTATGATGATATTTTAAATGAAGCAAGGCCATATTTTACAAATGA	5568	
QY	4429	CAGTCAACTAGTGCACATTTGACATGGATGAAGACCTGACACTATCTCTGGCTACTTTCAA	4488						

QYATYTRARAKPDSLEVLTLNLPNGTYEIKVAENNAGIVFSDPFLFOTARSAPGV
VNLTVFALNSAVNLWLPNGKITSFKLSVKHARSIGVVKVDSLEVEDILASGL
PECNENSEFLSWSTSTPGLRVTPTVTTQSSSTAARSKISVWKEPISFWPTHLR
PYTTLFEVASTTEAGYIDSTIVTESVPEGPQNCIGNVTKAFSISWPTTIV
TKGFSYRVELYSGRILDSNKTDLRFAFTHLTPFTMDYVVAETSAVGPKSNLSV
FPTPVPVAGFDLQIAEVAETIRITWRKPRONGIIISOYRVKVSLETGVLENTLL
TQODESISNMPSPPEIMNLVPMIGRYEGSGEMSDLHSPASFIYNHPNDPASTRA
EQSSPVVTRNQYMTDITAQLSVVVRVLPVTEHTISVAFPTIMGEPPVTLVTR
RQVSPSSIQIINKNISSSILLYWDPPIYPMKTHIYIATFELDTNRAFQMTVDN
SLFITGLKTYIKORVAATVHGSSSEENDI FVRTPEDEPSPQDVQVTSVPS
ELRLKWSPEKNGIIIAEYVLAADTLFVNKTSTDIISDLKPYTLNISRSYT
ELRGHNOSSLLSVTSETPVPSAPENIYKNISGEIISFLPPRSPNGIIIOKTYIY
LKRNSHEARTINTTSLTOTIGGLKXYTHVLEVASATLKGRISSRPIISLTEDAP
DPPQNFVKQLSGVTVMLSWOPPLEPNGIILYYTVVWDSLSLRINATEASLVSD
LDYNDYGCAVTASTRFGDNKARSSINRTPEGSDPPNDVHYVNLSSSIIILFW
PVPKNGIIIOYVSYYNTSGTFVQNFLLQVTKESDNTVVSARIYLAIFSYTFWL
TASTSVGNKSDIIPVYTDQIDPEGVGNLTFSISSTAIHVSWEPPSQNGLVFY
YLSLNUQQSPPHMIPLVYENSIDFDDLEKYTVI FKITPSTKGFSEVTTOLHI
KTEODVPDTPPIINTFKNISSTISLLSWDPPPLKNGAILGYHLTLQGPRIANITFVTS
NHLVLELSPFTLYSFPAAATWKLGPSSILFFYDESAPLAPPQNLTLINYSDPV
MLTWSPLPGGLGVKVISFKIHEHETDVIYKNISGLQTDALGLEPVSITYSVSA
FTKVGNGNOYSNVEFTTQESVAVRNEECVARDWQSVSVRMDPRKNTGIIHYMI
PVGNSKTVSPDRPTFTFKLLPNTSVFEVRASTASAGESNRCDISTLPETHVPSAP
TNVAFNSVOSTGATITFKRPDIIFGYFONYKITOLRAQCREWPEECIEHQKQYL
YEANOTEEVTHGLKFERWYRFQVAASTNVGYSNASEMI STQTLPGPDPGPPNVHVA
TSFPGINISWSPAVITGTFYLLDVKSVDDDPNISFLKSNEENKTYIINLVETR
YSVITAFVGNVSATVDSKSAEVIITLESVPKDPNNMTFQKIDEVTKFQTLFL
PSPQNGNIRVYQALVREDDTAVQIHFNSIIQKTDTSIIAMLEGLKGGHYNISVY
AINASAGVQKQWIRITMDIKAPAPKSKPIPIRDATGKLLVSTTIRIMPCTKYSND
HGPRIINVQVLAETGAQOQGNVTKWDAYFNKARPYFTNEGPPNPPCIEBKTKFSND
ELVIGADNACMIRGNEEKICNGPLKPKQYLFKPRATNVMGQFTDSYSDDPIKTLGE
GLISERVEILLSTCILSIIILLGTAFAPFVIRQKQEGGTYSPRDAELIITKSKLD
QLITVADLEKRLLSYKRSKIPISKSKFLQHVSELCTNSNLFQEFSESLPKF
LDLSSTDLADLPNKAARWFTPNKIPYNNRNVKLIADVSLPGSDYINASVSYGLCPNE
FIATQPLGTFQDFRVMVWFTKTLVMLTQCFEKGRIRCHQYWPEDNKPVTVFGDI
VITKLMEIDIQWITRDLKIERHGDCTVRCQNFQGWPEHGVPEMTPLIIRKVLVRT
SRAHDTTPMVHSCAGVGTGFIADLHDLTQHNHDFVDIYGLVAELRSEKVMQVN
LAQYIFLHCILDLNLSKGGHQVCFVNTSLQMDSLDAMESGDVELEWEETM"
177. .230
/genes="PTPRQ"
231. .7082
/genes="PTPRQ"
/product="glomerular mesangial cell receptor
protein-tyrosine phosphatase"
342. .638
/genes="PTPRQ"
/notes="fibronectin type III-like domain"
639. .1088
/genes="PTPRQ"
/note="fibronectin type III-like domain"
1089. .1352
/genes="PTPRQ"
/notes="fibronectin type III-like domain"
1353. .1622
/genes="PTPRQ"
/note="fibronectin type III-like domain"
1623. .1868
/genes="PTPRQ"
/notes="fibronectin type III-like domain"
1869. .2144
/genes="PTPRQ"
/note="fibronectin type III-like domain"
2145. .2450
/genes="PTPRQ"
/notes="fibronectin type III-like domain"
2451. .2732
/genes="PTPRQ"
/note="fibronectin type III-like domain"
2733. .3014
/genes="PTPRQ"
/notes="fibronectin type III-like domain"
3015. .3329
/genes="PTPRQ"

sig_peptide

mat_peptide

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

/note="fibronectin type III-like domain"
3330. .3623
/genes="PTPRQ"
/note="fibronectin type III-like domain"
3624. .3899
/genes="PTPRQ"
/note="fibronectin type III-like domain"
3900. .4184
/genes="PTPRQ"
/note="fibronectin type III-like domain"
4185. .4523
/genes="PTPRQ"
/note="fibronectin type III-like domain"
4524. .4787
/genes="PTPRQ"
/note="fibronectin type III-like domain"
4788. .5099
/genes="PTPRQ"
/note="fibronectin type III-like domain"
5100. .5399
/genes="PTPRQ"
/note="fibronectin type III-like domain"
5400. .5864
/genes="PTPRQ"
/note="fibronectin type III-like domain"
5901. .5969
/genes="PTPRQ"
/note="transmembrane-region site"
6129. .6950
/genes="PTPRQ"
/note="protein-tyrosine phosphatase domain"

ORIGIN

Query Match 74.5%; Score 5294.2; DB 10; Length 7396;

Best Local Similarity 84.7%; Pred. No. 0;

Matches 6048; Conservative 0; Mismatches 1049; Indels 45; Gaps 8;

QY 1 TAAATTGTGTACTTGCAGAGGATCTGCTTTTAAATCAATTAATGACGGCAACATTTCCT 60

Db 74 TAAATTGTGTACTTGCAGGGGATCTGACTTTTAAATCAATTAATGACGGCAACATTTCCT 133

QY 61 CTAGAGCCATCAATGTGATTTACTGGCTG-AAAAATGTAATAAGATGGATTTCTTAT 119

Db 134 CCAGAGCCATCAATGTGATTTGCTGTAGAAAAAATGTAATAAGATGGATTTCTTAT 193

QY 120 CATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTGATGTTCCATGTCGT 179

Db 194 CTCCTTCTCTTCTCTCTCATTTGGAACCTTCAGAGTCACAGGTGATGTTTCCAGTTCTTT 253

QY 180 TCCTGTACTAGTAGATATAACCATCTCTTCAATTTCT---ACAACATACACCTCACC 236

Db 254 TGATGGCATCGGTATGATATACCGTATCTTCAGTTCTGCCACACCTACAGCTCGCC 313

QY 237 TGTACTAGATAGTGACACCAAAATGTAAACAAACCCAGGGCCTTCCAGTCTTCTAGCCGG 296

Db 314 TGTGATAGAACACTGGCAACAAATGTGACAAACCCGGGSCCTTCCAGTCTTCTAGCCGG 373

QY 297 GGAAAGAGTCGAGATCTGTGGGATTTCTTGTCTTGGAAATACACCACTTAATCCAAATGG 356

Db 374 GGAGAGAGTTGGCTCTGTCTGGGATACTCTCTTCTGTGGAAATACCCCTCTTAACCAATGG 433

QY 357 AAGGATATATCTTACATTTGTCAAATATAGGAAGTTTCTCGTGGATGCAACACATATA 416

Db 434 AAGAATCATATCTTATGTTTGAAGTACAGGAAGTCTGTCCATGGATGACGACGACATA 493

QY 417 TACACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTCTTCTTAAATCTTAAATCCTGG 476

Db 494 TACCCGAGCGAGAGGAGCCAGACAGTCTGGAGGTCTTCTTCACTAACTTAATCCTGG 553

QY 477 AACACATATGAATTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536

Db 554 AACACATATGAATTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 613

QY 537 TCCATTTCTCTCCAACTCAGAAAGTGTCCAGGAAAGTGGTGAATCTCAAGTTGA 596
DB 614 CCCATTTCTTTTCAAACCTGAGAAAGTGTCCGGGTAAAGTAGTTAACTCAAGTGA 673
QY 597 GGGCTACAAAGCTTTCAGCAGTAAAGTGTATTTGATATTTTAACTCCGGCAACAAATGGCAA 656
DB 674 AGCTCTCAAACCTACTCTGCAGTAAACTTGAATTTGGTATTTTAACTCCGGCAACAAATGGCAA 733
QY 657 AATTACCAAGCTTCAAGANTAGTGTCAAACATGCGCAAGTGGGATAGTAGTGAAGATGT 716
DB 734 AATCACCAAGCTTCAAGANTAGTGTCAAACATGCGCAAGTGGGATAGTAGTGAAGATGT 793
QY 717 CTCNAATCAGAGTGAAGACATTTTGTCTGGGAATTTCCCAAGATGCAATGAGANTAGTGA 776
DB 794 CTCCTCAGAGTGAAGACATTTTGTCTGGGAATTTTAACTGNAATGTAATGAGAACAGCA 853
QY 777 ATCTTTTATGAGTACAGCAGCCTCTCTCAACCTTGGTAGAGTTACACTCCATC 836
DB 854 ATCTTTTGTGGATACCAAGCCTCTCTCAACCTGGTAGAGTGACACTACGT 913
QY 837 GCGTACCAACATTTCAATCAAGCAGCTTTGACACAGAAATGAGATCAGCTCTGTGTGGAAGA 896
DB 914 TCGCACCAACAGCTCATCAGCAGCGCGGCAAGAGTAAAGTCAAGCTCTGTGTGGAAGA 973
QY 897 GCCTATCAGTTTGTAGTGAACACTTGAAGACCTTATACAAATATCTTTTGAAGTTTC 956
DB 974 GCCCATCAGTTTGTGTGAGTCACTCTTGAAGACCTTATACAAACGTATCTCTTTGAAGTCTC 1033
QY 957 AGCTGTACAACTGAAAGCAGTTATATGATAGTACGATTTGTGAGACACCAAGATCAGT 1016
DB 1034 AGCGTTACACGGAAGCAGATATATTTGACAGCAATTTGTAGAGCGCCAGATCAGT 1093
QY 1017 GCCTGAAGGACCAACCAAACTCGGTAAACAGGCAACATCAAGGAAAGTCTTTTCAAT 1076
DB 1094 GCCTGAAGGACCAACCAAACTGTATATGCGGAATGTCAAGGAAAGGCTTTTCAAT 1153
QY 1077 TTTATGGGACCAACCAATATAGTAAACAGGAAATTTAGTTATAGAGTTGAATATATGG 1136
DB 1154 TTCTGGGACCAACCAATATTTGTCAAGGAAAGTTTGTATACAGAGTTGAATTTGTATGG 1213
QY 1137 ACCATCAGTTCGATTTTGGATACAGCAGCAAGACCTCAAGTTTGCATTTCACTAACCT 1196
DB 1214 GCATCAGGCGGATTTGGATATAGCAGAAAGACCTCCGGTTTTCGTTTACATCT 1273
QY 1197 AACACATTTACATATGATGATGTCTATATTTGGGCTGAAACAGTGCAGGGACTGGGC 1256
DB 1274 CACACCATTCACAAATGTACGATGTCTATGTCTCGGGAACCCAGTGTGGGTTGAC 1333
QY 1257 CAAGTCAAAATTTTCACTCTCAACAGATTTCCAGGGGAGTGTTCATTTACA 1316
DB 1334 CAAGTCAAAATCTTTCAATTTTACTCTCCCAAGATTTTCCAGGAGCAGTGTTCATTTACA 1393
QY 1317 ACTTCAGAGGTAGATCCAGCAAGTAAAGATTTCTGGAGAAACCAAGCAACCAAA 1376
DB 1394 GATTCGAGAGGTAGAGCCACAGAAATAGAAATTTACTTGGAGAAACCTCGAACCAAA 1453
QY 1377 TGGAAATTTTAAACCAATACCGAGTGAAGTGTAGTTTCCAGAGACAGGAATATTTTGA 1436
DB 1454 CGGAATCATCAGCCAGTACCGTGTGAAGTGTCAAGTCTTACAGACAGGAGTAGTTTGA 1513
QY 1437 AAATATCTTGTCTCACTGGAAATATGATATATAAATGACCCCATGGCTCCAGAAATGT 1496
DB 1514 AAATATCTTGTCTCAAGGACAAAGTGTCTAAGCAACCCCATGTCTCCAGAAATAT 1573
QY 1497 GAACATAGTAGCCCAATGTTAGATATATAGGGTTACGAGGATGTCGTCTGACCT 1556
DB 1574 GAACCTAGTAGACCCCATGATAGGTTCTATGAGGCTCAGGAGAAATGTCTTCGACCT 1633
QY 1557 TCACTCACTTGTCTACATTTATATAAAGCCCATCAGATATAAAATTTTCTCTCAGGAA 1616
DB 1634 GCACTCCCGGCTTCAATTCATATACAAAGCCCATCACAATGATCTTCTCGCAGCAC 1693
QY 1617 TAGAGCTGAAGACCAAGTCTTCAAGTTGTAACTACAAGGAATCAGTATATTAATCTGACAT 1676

DB 1694 TAGAGCTGAAGCAAGTTTCAAGTTGTAAACCAAGGAATCAGTATATATGACCGCAT 1753
QY 1677 TGCAGCTGAACAGCTGTCTTATCTTATCAGGAGACTTGTACCTTTTCACTGAGCACAATGAT 1736
DB 1754 CACTGCCGAACAGCTGTCTTACGTGTCTCAGCGGCTTGTACCAATCTCACTGAGCACAATGAT 1813
QY 1737 TAGGTATCTGTCTTTCACCAATCATGGGAGAGGACCAACCAAGTGTCTCAGTGTAGGAC 1796
DB 1814 TAGCGTATCTGTCTTTCACCAATCATGGGCGAAGGACCAACCAAGTGTCTTCTGTTAGGAC 1873
QY 1797 ACCTCAGCAGTGCAGCTTCAATTAATAACTATAAAATATAAAATATTAAGTTCTTCATC 1856
DB 1874 ACCTGAGCAGTGCAGCTTCAATTAATAACTATAAAATATAAAATATTAAGTTCTTCATC 1933
QY 1857 TATTTTGTATATTTGGGATCTCCAGATATCCCAATGGAATAATACTCACTATACGAT 1916
DB 1934 TATTTTACTGTATTTGGATCTCCAGATATCCCAATGGAAGATCACTCATTTACAGAT 1993
QY 1917 TTAGCAATGGAATTCGATACAAACAGAGCATTTCCAGATAAATCAATAGATAACAGCTT 1976
DB 1994 CTACGCAACGAGCTGGAACAAACAGAGCATTTCCAGATGACCACTGTAGATAACAGCTT 2053
QY 1977 TCTCATAAACAGGTTTAAAGAAATACAAATAACAAATGAGAGTGGCAGCTCAACCCA 2036
DB 2054 TCTCATACAGGTTTGAAGAAATATACAGATACAAAATGAGGTTGTGCTCAACTCA 2113
QY 2037 CGATGAGAGAAAGTTCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTTCAAGAGATGA 2096
DB 2114 TGTGTGGGAAAGTTCTTTGTCTGAGAAATGACATCTTTGTGAGAACTTTCAAGAGATGA 2173
QY 2097 ACCGGAATCATCTCAAGATGTCCAGATTAATGTATGTTTACCGAGATGAAATAAGGTT 2156
DB 2174 GCCAGATCATCACCTCAAGATGTCCAGTAACTGTGTCTTCCAGTGAACCTGAGAT 2233
QY 2157 GAAGTGTCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGAGTGTCTATATA 2216
DB 2234 GAAGTGTCCCGCTCGAAAGAACCTTAAAGGATCATTTATGCTTATGAGTGTCTTACCA 2293
QY 2217 AAATAAGATATCTTTATATATGAGAAACATCAACAAACAGACATAATATAAGGAAT 2276
DB 2294 AAATGACAGACCTTGTGTTTGAAGAAACCTCAACCAAGACATCATTAATAGTACCT 2353
QY 2277 AAGACCTCAACCTCTATTAACATTTCTGTAAAGTCTTACACAGATTTGGTCAATGCAA 2336
DB 2354 GAAACCTTACACATCTATAATATTTCTATCGGTCTGTATACAGACTTGGCCATGGCAA 2413
QY 2337 TCAGGTATCTTCTTACTCTCTGTAAAGACTTGGGAGACTGTGCTGTATAGTGCACAGA 2396
DB 2414 TCAGTCACTCTCATTTGCTCTGTAAAGACTTTCAGAGACTGTGCTGTATGACACAGA 2473
QY 2397 AAATATCACTTCAAAATAATTTTCTTCGAGAGATTTGAGCTATCATTTCTTCCCAAG 2456
DB 2474 AAATATCTTCAAAATAATTTTCTTCAGAGAGATTTGAGATATCTTCTTCCCAAG 2533
QY 2457 TAGTCCCAATGGAATCATTAATAATAATATATTTATCTCAAGAGAAAGTAAATGGAATGA 2516
DB 2534 GAGTCCCAACGCAATCATACAAATAACAAATTTTACCTCAAGAGAAAGTAAATAGCCATGA 2593
QY 2517 GGAAGAACTATATAATACAACTCTTTTAAACCAAAACATTAAGTACTGAAGAAATATAC 2576
DB 2594 GGCAGAACTATAATAATACAACTCTTTTGAACCAACCAATTTGGAGGACTGAAAAAGTATAC 2653
QY 2577 CCATATATCATTTGAGTGTCTGTAGTACACTTGAAGGTGAGGAGTTCCGAGTGTCTCC 2636
DB 2654 CCATATATGTAATCGAGTGTCTGTAGTACACTTGAAGGTGAGGAGTTCCGAGTGTCTCC 2713
QY 2637 CATAGTATACGAGGAGAAAGTGTCTCTGATTTCTCCCTCAAGACTTCTCTGTAAA 2696
DB 2714 AATAGCATATTTGACCGAGGAGAGCGCTCTGATTTCTCTCTCAGAACTTCTCTGTGAA 2773
QY 2697 ACAGTGTCTGTGTCAAGGTGAGTGTGTCAATGGCAACCAACCTCTGAGGCAATGGAAT 2756

Db 2774 ACAGTTGTCTGGTGTCAACGGTGTATGTGTCTATGGCAACCAACCCCTTGGAGCCAAATGGAAT 2833
Qy 2757 TATCCTTTTATTACAGAGTTTATGTCTGGATAGATCATCATTTAAAACCTATTAATGTCTAC 2816
Db 2834 TATCCTCTTATTACAGAGTTTATGTCTGGATATAAATCATCTTAAGAGCTATTAATGCAAC 2893
Qy 2817 TGAACATCATTTGGAGTTATCAGATTTGGATTATTAATGTGGAATACAGTCTTTATGTAAC 2876
Db 2894 TGAAGCATCACTGGTGTGTCTGACTTAGACTATAATGTGACTACGGTCCCTGTGTAAC 2953
Qy 2877 AGCTAGCACAGATTTGGTGTGATGGGAAACAGGAACAATATCATTTAGCTTTCAACACC 2936
Db 2954 AGCTAGCACAGATTTGGCGATGGGAACGCAAGACGAGCATCATTAACCTTTGCAACACC 3013
Qy 2937 AGGGAGACCAACGAGCTCTCCCAAGATGTTTATTATGCAAACTCAGTCTTCTATC 2996
Db 3014 AGAGGAGAGCCAGTGATCTCCCAACGATGCTCACTATGTGAATCTCAGTTCCTCATC 3073
Qy 2997 AATAAATCTTTTCTGGACACCTCTTTCAAAACCTAATGGATTTATACAATATTACTCTGT 3056
Db 3074 GATTAATCTTTTGGACTCTCTCTGTAAACCTAATGGATTCATTAATATTCTGT 3133
Qy 3057 TTATTAACAGAAATPACTTCAGGTACTTTTATGCAAGAAATTTTACATCCATGAACTAACCA 3116
Db 3134 TTATTAACAAATPACTTCAGGTACTTTTGTGCAAGAACTTTACACTCTTCAAGTAACCAA 3193
Qy 3117 TGACTTTGACATATGACTGTATCCACAATATTAGATAAATGACAAATATTGACTACTA 3176
Db 3194 GGAGTCTGACAAATGTGACAGTGTCTGCAAGGATATACAGACTGGCAATATTGACTACTA 3253
Qy 3177 TACATTTTGGTTAAACAGCAAGTACTTTCAGTTGGAAATGGGAATPAAAGCAGGTGACATCAT 3236
Db 3254 TACATTTTGGCTTAACAGCAAGTACTTTCAGTTGGAAATGGGAATPAAAGCAGGTGACATCAT 3313
Qy 3237 TGAAGTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAACCTGACTTACGAATC 3296
Db 3314 TCATGTATACACAGACCAAGACATACCTGAAGTCTCTGTTGGAACCTGACTTTTGAGTC 3373
Qy 3297 CATTTGCTGACCTGCAATTAATGTAAAGCTGGTCCACCGGCTCAACCAACCGGCTAGT 3356
Db 3374 CATTTTCACTGCAATTCATGTGAGCTGGGAGCCACCATCCCAACCAATGGCCGTGT 3433
Qy 3357 CTTCTACTATGTTTCACTGATCTTTACAGACAGA--CTCCTCGCATGTGAGACCACTCT 3413
Db 3434 CTTCTACTACTTTTCACTGATTTACGCAAGTCCCCACCGCACATGATACCACTCT 3493
Qy 3414 TGTACATATGAGAGAGCATATTTTGTATATCTGGAATAATACACTGATTTATATTT 3473
Db 3494 GGTACCTATGAAACACAGCATAGACTTTGATGATCTGGAATAATATACCTGACTATATTT 3553
Qy 3474 AAAAATTACTCCATCAACAGAAAGGATTCCTGATACCTATACCTGCTCCAGCTATACAT 3533
Db 3554 TAAATCACTCCGTCACAGAAAGGATTTCTCTGAGACCTATPACTACCCAACTACACAT 3613
Qy 3534 CAAGACTGAAGAGATGTCACAGAACTTCACCAATAATCAACACTTTTAAAAACCTTTC 3593
Db 3614 CAACAGAGAGAGAGCTTCAGACACTCCACCAATATCAACACTTTTAAAAATCTTTC 3673
Qy 3594 CTTACCTCAGTTCTCTTATCATGGGATCCCCAGATAAAGCCAAATGGTGCAATAAAG 3653
Db 3674 TTCTACCTCATTTCTTATCATGGGATCCCCCATTTAAAGCCAAACGGTGCAATCTAGG 3733
Qy 3654 TTATGATTTTAACTTTTACAGGACCAATGAAATTTATCTTTCAATTTACTTTCTGATTAATTA 3713
Db 3734 TTACCAATTTAACTTTTACAGGACCCCAAGCAACCATACCTTTTCTGCTTCCGCTAACCA 3793
Qy 3714 CATATAATTTGGAAGACTTTTCACTTTTACATTTATATAGCTTTTCTGCGCCAGAAC 3773
Db 3794 CATAGTACTGGAAGAACTCTCGCAATTTACCTTATATAGTTTTTTTCTGCTGCAAGAAC 3853
Qy 3774 TAGAAAAGGACTTGTGCTTCCAGTATCTTTTCTTTTACACAGATGAGTCAAGTCCGCTT 3833
Db 3854 AATGAAAGGACTCGGTCCTTCCAGTATCTTTTCTTTTACACAGATGAGTCAAGTCCGCTT 3913

Qy 3834 AGCACTCCACAAAATTTGACTTTTAACTCAACTGTACTTTCAGACTTTTGTATGCTGAAATG 3893
Db 3914 AGCACTCCCAAGATTTGACTTTTAACTCAACTGTACTTTCAGACTTTTGTATGCTGACATG 3973
Qy 3894 GAGCCCAAGTCTCTTTCAGAGTGTATTGTTAAAGTATATAGTTTAAATTTTCAATGCA 3953
Db 3974 GAGCCCAAGTCTCTTTCAGAGTGTATTGTTAAAGTATATAGTTTAAATTTTCAATGCA 4033
Qy 3954 TGAACCTGACATATATATTATAAGAAATATATCAGGATTTAAACCTGAAGCCAACTTGT 4013
Db 4034 TGAGACTGATATCTGATTTTATAAGAAATATTTTCAAGTCTTTCAGACTGATGCAAACTTGA 4093
Qy 4014 TGAGACTGGAACCAAGTCAAGCCTCTCTATCCGTGTATCTGCTGCTTCAACCAAGTTGAAA 4073
Db 4094 GGGACTGGAGCTGTGTAGCACCCTCTCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 4153
Qy 4074 TGGCAATCAATTTAGTAAATTTTCAAAATTTCAAAACCCCAAGAAATCAGTTTCCAGATGCTGT 4133
Db 4154 TGGCAATCAGTATAGTAAATTTTCAAAATTTTCAAAACCCCAAGAAATCAGTTTCCAGAGCTGT 4213
Qy 4134 GAGAAATGAGTGGATGGCAACTAGCTGGCAGTCAAGTTTGTAGTGAATGGATCCACC 4193
Db 4214 ACGGAATATAGAGTGGTGGCACTGACTGGCAGTCAAGTATCGGTGAGGTGGGACCCACC 4273
Qy 4194 CAABAAGGCAATGGAATAATAACGAGTATATGTTAAAGGAAATTTCTACAAA 4253
Db 4274 GAGAAAAAATAATGGAATAATTTTCACTATATATGTAACGGTTGGAGGAAATTTTACCAA 4333
Qy 4254 AGTTTCTCCCAAGATCACTGTACATCTTTCATAAAGCTTTTGTCCAAATACCTCATATGT 4313
Db 4334 AGTCTCTCCCAAGACCCCAAGTCAAGCTTTCACAAAATCTTCTTCCCAATACGTCATACGT 4393
Qy 4314 CTTTAAAGTAAAGCTTTCAACCTCAGCTGGTGAAGGTGATGAAGAACACATGCCATGTGAC 4373
Db 4394 CTTTGAAGTAAAGGCTCAACATCAGCGCGCAAGGCAATGAAAGCAGGTGTGACATCAG 4453
Qy 4374 CACACTCTCGAAGACAGTTCCAGTGTCCCAAAATATGCTTTTCTGATGTTTCAGTC 4433
Db 4454 CACACTCCCGAGAGACAGTTCCAGTGTCCCAAAATATGAGCGTTTCTTAATGTCAGTC 4513
Qy 4434 AACTAGTGAACCAATTTGACATGGAATGAAGCTGACATCTCTTGGCTACTTTTCAAAATTA 4493
Db 4514 AACTAGTGAACCACTAAACATGACAAAACCTGATCTATCTTTGGATATCTTCCAAAATTA 4573
Qy 4494 GAAATTTACCATCAACTTCTGCTCAAAAATGCAAAAGATGGGAATCCGAGAAATGTGT 4553
Db 4574 CAAGATAACTACTCAACTTCTGCTCAGAAAGTGCAGAGAAATGGAACTGAAAGATGTAT 4633
Qy 4554 TGAATATCAAAAAATTTCAATACCTCTATGAAGCTCACTTAACTGAAGACACAGTATATGG 4613
Db 4634 TGAACATCAGAAAGATCAATACCTCTATGAGCAACCAACCAAGAAAGACAGTGCATGG 4693
Qy 4614 ATTAAAGAAATTTAGATGGTATAGATTTCAAGTGGCTGCCAGCAACCAATGCTGGCTATGG 4673
Db 4694 GTTAAAGAGTTTCGATGGTATAGATTTCCAGTGGCTGTAGCACCATGTTGGGTACAG 4753
Qy 4674 CAATGCTTCAAACTGGAATTTTCAAAAACCTGCTGCTGCCCTCCAGATGGCTCTCTGA 4733
Db 4754 TAATGCTCGGAATGGATTTTACACAAACCGTGCCTCCCTCTCTGACGGCTCTCTGA 4813
Qy 4734 AATGTTTCATGATAGTGAACCATCACTTTTATAGCATCAGATAAGCTGAGTGAACCTGC 4793
Db 4814 AATGTCACGTTGGTGGCAACCATCACTTTTGTATCAATATCAGCTGGAGTGAACCTGC 4873
Qy 4794 TGTCAATTTACTGGAACCAACATGTTTATCTGATTTGATGTCAAAATCGGTAGATAATGATGAA 4853
Db 4874 TGTCAATTTACTGGAACCAACATTTCTACCTGATTTGATGTCAAAATCGGTAGATGATGATGAT 4933
Qy 4854 TAAATATCTCTTTCATCAAGTCAAAATGGAAGAAATTAACCAATAGAAATTAAGATTTAGA 4913
Db 4934 TAAATATTTCTTCTTCTGAAATCAATGGAAGAAACCAAAACCAAGAGATTTAAACACTTGA 4993


```
Qy 6054 TGAAGAGCTTTGCAAAACAAACCTAAAGTTTCAAGAAGAAATTTCCGAATATACAAA 6113
Db 1773 TGAAGAGCTTTGCAAAACAAACCTAAAGTTTCAAGAAGAAATTTCCGAATATACAAA 1832
Qy 6114 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGCTGCTTGGAAATAGACAAACACCG 6173
Db 1833 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGCTTGGAAATAGACAAACACCG 1892
Qy 6174 TTTCCCAACATAAAACCATATAATAATAACAGAGTAAGCTGATAGCTGACGCTAG 6233
Db 1893 CTTCCCAACATAAAACCAT--ATAATAATAACAGAGTAAGCTGATAGCTGACGCTAG 1949
Qy 6234 TGTTCAGGTTCCGATTAATAATAGCAGCTATAATTTCTGGTTATTTATGTCACAAATGA 6293
Db 1950 TGTTCAGGTTCCGATTAATAATAGCAGCTATAATTTCTGGTTATTTATGTCACAAATGA 2009
Qy 6294 ATTTATTTGCTACTCAAGSTCCACTACCAGAGAACAGTTGGAGATTTTGGAGATGGTGTG 6353
Db 2010 ATTTATTTGCTACTCAAGSTCCACTACCAGAGAACAGTTGGAGATTTTGGAGATGGTGTG 2069
Qy 6354 GGAACACAGGGCAAAACATTTAGTAATGCTAAACACAGTGTTTTGAACAGGACGGATCAG 6413
Db 2070 GGAACACAGGACAAACATTTAGTAATGCTAAACACAGTGTTTTGAACAGGACGGATCAG 2129
Qy 6414 ATGCCATCAGTATTTGGCCAGAGGACAAAGCCAGTTACTGTCTTTTGGAGATATAGTAT 6473
Db 2130 ATGCCATCAGTATTTGGCCAGAGGACAAAGCCAGTTACTGTCTTTTGGAGATATAGTAT 2189
Qy 6474 TACAAAGCTAATGGAGATCTTCAATAGATATGAGTATGAGTATCAGGATCTGAAATTTGAAG 6533
Db 2190 TACAAAGCTAATGGAGATCTTCAATAGATATGAGTATCAGGATCTGAAATTTGAAG 2249
Qy 6534 GCATGGGATTTGATGATCTGTTGCACAGTGTAACTTTACTGCTGCGCAGAGCATGGGGT 6593
Db 2250 GCATGGGATTTGATGATCTGTTGCACAGTGTAACTTTACTGCTGCGCAGAGCATGGGGT 2309
Qy 6594 TCTGTAGAACAGGCGCCCTCTAATTTCACTTTGTGAAGTTGGTTGGAAGCAAGCGGACCA 6653
Db 2310 TCTGTAGAACAGGCGCCCTCTAATTTCACTTTGTGAAGTTGGTTGGAAGCAAGCGGACCA 2369
Qy 6654 TGCACACACCTATGATTTGTTCACTGAGTGTGAGTGTGGAAGCAAGCTGGAGTTTAT 6713
Db 2370 TGCACACACCTATGATTTGTTCACTGAGTGTGAGTGTGGAAGCAAGCTGGAGTTTAT 2429
Qy 6714 TGCTCTGGACATTTAACAACAATATAAATGACCATGATTTTGTGATATATATGACT 6773
Db 2430 TGCTCTGGACATTTAACAACAATATAAATGACCATGATTTTGTGATATATATGACT 2489
Qy 6774 AGTAGCTGAACGTGAAGTGAAGAAATGTGCATGGTGCAGAAATCTGGCAGATATATCTT 6833
Db 2490 AGTAGCTGAACGTGAAGTGAAGAAATGTGCATGGTGCAGAAATCTGGCAGATATATCTT 2549
Qy 6834 TTTACACAGTGTGATCTCTATCAATTAAGGAAATTAATCAGGCCATCTGTTT 6893
Db 2550 TTTACACAGTGTGATCTCTATCAATTAAGGAAATTAATCAGGCCATCTGTTT 2609
Qy 6894 TGTAACTATTCAGCACTTCAGAAAGTGGACTCTTTGGAGCCCATCGAAGTGTATGTA 6953
Db 2610 TGTAACTATTCAGCACTTCAGAAAGTGGACTCTTTGGAGCCCATCGAAGTGTATGTA 2669
Qy 6954 GCTTGAATGGAAAGAAACCACTATGTAATATTTACAGACCAAGGATACAAATTTGGAAGAGA 7013
Db 2670 GCTTGAATGGAAAGAAACCACTATGTAATATTTACAGACCAAGGATACAAATTTGGAAGAGA 2729
Qy 7014 TTTTAAATCCAGGGGCCAAAGTTACCCCTCATTTCTCCGAATTTGAATGTCACCT 7073
Db 2730 TTTTAAATCCAGGGGCCAAAGTTACCCCTCATTTCTCCGAATTTGAATGTCACCT 2789
Qy 7074 TAAAGAAATATCTATGCTCTCTCCTCCTGCTGCTT 7108
Db 2790 TAAAGAAATATCTATGCTCTCTCCTCCTGCTGCTT 2824
```

```
RESULT 9
LOCUS 132039
DEFINITION Sequence 21 from patent US 5585233.
ACCESSION 132039
VERSION 132039.1 GI:1822830
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3973)
AUTHORS Moller,N.P.H., Moller,K.B. and Ullrich,A.
TITLE PRP-S31: A novel protein tyrosine phosphatase
JOURNAL Patent: US 5585233-A 21 17-DEC-1996;
FEATURES
    Location/Qualifiers
        1..3973
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN
Query Match 36.8%; Score 2618.6; DB 6; Length 3973;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;
Qy 4386 AACAGTTCCAGTGTTCACCAAAATATTCCTTTTCTGATGTTTCAGTCAAAGTCAAC 4445
Db 33 AAAAGTTCCAGTGTTCACCAAAATATTCCTTTTCTGATGTTTCAGTCAAAGTCAAC 92
Qy 4446 ATTGACATGATTAAGACCTGACACTATCTTGGCTACTTTCAAAATTTACAAATTACAC 4505
Db 93 ATTGACATGATTAAGACCTGACACTATCTTGGCTACTTTCAAAATTTACAAATTACAC 152
Qy 4506 TCAACTTCGTGCTCAAAATTCGAAAGATCGGAATGTTTGAATATCAAAA 4565
Db 153 TCAACTTCGTGCTCAAAATTCGAAAGATCGGAATGTTTGAATATCAAAA 212
Qy 4566 AATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGATTAAGAAAT 4625
Db 213 AATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGATTAAGAAAT 272
Qy 4626 TAGATGATATAGATTTCCAGTGGCTGCCAGCACCATCTGGCTATGGCAATGCTTCAA 4685
Db 273 TAGATGATATAGATTTCCAGTGGCTGCCAGCACCATCTGGCTATGGCAATGCTTCAA 332
Qy 4686 CTGGATTTCTACAAAACCTCTGCTGCGCTCCAGATGGTCTCTCGAAATGTTCAATG 4745
Db 333 CTGGATTTCTACAAAACCTCTGCTGCGCTCCAGATGGTCTCTCGAAATGTTCAATG 392
Qy 4746 AGTAGCAACATCACTTTTACGATCAGCATAAAGCTGGAGTGAACTGCTGCTCATTCGG 4805
Db 393 AGTAGCAACATCACTTTTACGATCAGCATAAAGCTGGAGTGAACTGCTGCTCATTCGG 452
Qy 4806 ACCAAGTGTATCTGATTTGATGTCAAATCGGTAGATATAATGATGAATTTAATATATCT 4865
Db 453 ACCAAGTGTATCTGATTTGATGTCAAATCGGTAGATATAATGATGAATTTAATATATCT 512
Qy 4866 CATCAAGTCAAAATGAAGAAATTAACCATAGAAATTAAGATTTTGAATATTTACAAG 4925
Db 513 CATCAAGTCAAAATGAAGAAATTAACCATAGAAATTAAGATTTTGAATATTTACAAG 572
Qy 4926 GTATTCCTGTAGTATCACTGCAATTTTACTGGGAACATTTAGTGTGCATATGTAGAAGG 4985
Db 573 GTATTCCTGTAGTATCACTGCAATTTTACTGGGAACATTTAGTGTGCATATGTAGAAGG 632
Qy 4986 GTCAAGTGTGAATTAATGTTTACTACTCTTTAGAAATCAGGCCCAAGGACCACTAACAA 5045
Db 633 GTCAAGTGTGAATTAATGTTTACTACTCTTTAGAAATCAGGCCCAAGGACCACTAACAA 692
Qy 5046 CATGACATTTCAAGATACCAAGTGAAGTTACAAATTTCAATTAACGTTCTCTCTCTCC 5105
Db 693 CATGACATTTCAAGATACCAAGTGAAGTTACAAATTTCAATTAACGTTCTCTCTCTCC 752
```

QY 5106 TTCTCAACCTAATGGAATATCAAGTATATCAAGCTCTGGTTTACCGAAGATGATCC 5165
DB 753 TTCTCAACCTAATGGAATATCAAGTATATCAAGCTCTGGTTTACCGAAGATGATCC 812
QY 5166 TACTGTCTCGAGTTTCAAACTCAGTATATACAGAAAAACAACATTTGTCATGTC 5225
DB 813 TACTGTCTCGAGTTTCAAACTCAGTATATACAGAAAAACAACATTTGTCATGTC 872
QY 5226 AATGCTAGAGGACTAAAGGTGGACATACATCAATATCAGTGTGTTACGAGTCAATAG 5285
DB 873 AATGCTAGAGGACTAAAGGTGGACATACATCAATATCAGTGTGTTACGAGTCAATAG 932
QY 5286 TGTGCTGTCAGGTCCAAAGGTTCCGATGAGAAATCAACATGGATATCAAAAGCTCCAGCAG 5345
DB 933 TGTGCTGTCAGGTCCAAAGGTTCCGATGAGAAATCAACATGGATATCAAAAGCTCCAGCAG 992
QY 5346 ACCAAAAACCAACCCCTATTTATGATGCGCACAGGAAATCTGTTGAGCTTTCAAC 5405
DB 993 ACCAAAAACCAACCCCTATTTATGATGCGCACAGGAAATCTGTTGAGCTTTCAAC 1052
QY 5406 AACAAATTACAAATCAGAAATGCCAATATGTTACTACAGTGATGATCATGGACCAATAAAAA 5465
DB 1053 AACAAATTACAAATCAGAAATGCCAATATGTTACTACAGTGATGATCATGGACCAATAAAAA 1112
QY 5466 TGTACAAGTCTTTGCGACAGAAACAGGAGCTCAGCATGATGAGAAATGTAACAAAGTGGTA 5525
DB 1113 TGTACAAGTCTTTGCGACAGAAACAGGAGCTCAGCATGATGAGAAATGTAACAAAGTGGTA 1172
QY 5526 TGATGATATTTTAAAGCAAGGCCATATTTTAAAGAAAGGCTTTTCTTAACCTCTC 5585
DB 1173 TGATGATATTTTAAAGCAAGGCCATATTTTAAAGAAAGGCTTTTCTTAACCTCTC 1232
QY 5586 ATGACAGAGGAAGCAAGGTTTACTGCGAATGAGAAATCTCATCATAGTGGTCTGA 5645
DB 1233 ATGACAGAGGAAGCAAGGTTTACTGCGAATGAGAAATCTCATCATAGTGGTCTGA 1292
QY 5646 TAATGATGATGATTTCTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAGCAAA 5705
DB 1293 TAATGATGATGATTTCTGGCAATGAGCAAAATTTGCAATGGACCACTGAAGCAAA 1352
QY 5706 AAAGCAATATTTTAAATTTAGAGCTACAAATATATGGGCAATTTACTGACTCTGA 5765
DB 1353 AAAGCAATATTTTAAATTTAGAGCTACAAATATATGGGCAATTTACTGACTCTGA 1412
QY 5766 TTATTTCTGACCTGTTAAGACTTTAGGGAGGACCTTCAGAAAGACCTGAGATCAT 5825
DB 1413 TTATTTCTGACCTGTTAAGACTTTAGGGAGGACCTTCAGAAAGACCTGAGATCAT 1472
QY 5826 TCTTTCCGTCACTTTGTGATCTTTCAATAATTTCTCTTGGAAACAGCTATTTTGTGAT 5885
DB 1473 TCTTTCCGTCACTTTGTGATCTTTCAATAATTTCTCTTGGAAACAGCTATTTTGTGAT 1532
QY 5886 TGCAAGAAATTCGACAGAGCAGAAAGAGGTGGCAATATCTCTCTCAGGATGCAAAAT 5945
DB 1533 TGCAAGAAATTCGACAGAGCAGAAAGAGGTGGCAATATCTCTCTCAGGATGCAAAAT 1592
QY 5946 TATTGACATTAATTTGAGCTGATCAGCTCATCAGTGGCAGACTGGAAGTGAAGGA 6005
DB 1593 TATTGACATTAATTTGAGCTGATCAGCTCATCAGTGGCAGACTGGAAGTGAAGGA 1652
QY 6006 CGAGAGATTAACGCG----- 6020
DB 1653 CGAGAGATTAACGCGTACTCTTCATTTTTTTTAGACGCAAGAGATTTTTTGTGATCCA 1712
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTCTGCAACATGT 6053
DB 1713 GTTACTTAGTTATAGAAATCCATCAAGCCAAATAGCAAGAAATCTCTCTGCAACATGT 1772
QY 6054 TGAAGAGCTTTGCAACAAACCAACCTTAAGTTTCAAGAGAAATTTTGGAAATTCACAA 6113
DB 1773 TGAAGAGCTTTGCAACAAACCAACCTTAAGTTTCAAGAGAAATTTTGGAAATTCACAA 1832
QY 6114 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAACCG 6173

DB 1833 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGCTATGCTTGGAAATAGAGCAAAACCG 1892
QY 6174 TTTTCCCAACATAAAACCATATAATAATAACAGAGTAAAGCTGATAGCTAGAGCTAG 6233
DB 1893 CTTTCCCAACATAAAACCAT---ATAATAATAACAGAGTAAAGCTGATAGCTAGAGCTAG 1949
QY 6234 TGTTCAGGTTCCGGATTTATTAATGACCACTATATTTCTGTTTATTTATGTCCTCAATGA 6293
DB 1950 TGTTCAGGTTCCGGATTTATTAATGACCACTATATTTCTGTTTATTTATGTCCTCAATGA 2009
QY 6294 ATTTATTTGCTACTCAAGTTCCTACAGGAAACAGTTGGAGATTTTGGAGAAATGTTGTC 6353
DB 2010 ATTTATTTGCTACTCAAGTTCCTACAGGAAACAGTTGGAGATTTTGGAGAAATGTTGTC 2069
QY 6354 GGAAACCAAGGCGCAAAACATTAAGTAACTGTAACACAGTCTTTTGGAAAGGACGATCAG 6413
DB 2070 GGAAACCAAGGCGCAAAACATTAAGTAACTGTAACACAGTCTTTTGGAAAGGACGATCAG 2129
QY 6414 ATGCCATCAGTATTTGGCCAGAGGACAAACAGGCGATTAAGTCTTTTGGAGATATAGTAT 6473
DB 2130 ATGCCATCAGTATTTGGCCAGAGGACAAACAGGCGATTAAGTCTTTTGGAGATATAGTAT 2189
QY 6474 TACAAAGCTAATGAGGATGTTTCAATAGATTGGACTATCAGGATCTGAAATTCGAAG 5533
DB 2190 TACAAAGCTAATGAGGATGTTTCAATAGATTGGACTATCAGGATCTGAAATTCGAAG 2249
QY 6534 GCATGGGATTCATGACCTGTTTCCAGAGTAACTTTTACTGCTGCGCAGAGCATGGGT 6593
DB 2250 GCATGGGATTCATGACCTGTTTCCAGAGTAACTTTTACTGCTGCGCAGAGCATGGGT 2309
QY 6594 TCTGAGAAACAGCGCCCTCTAATTTCACTTTGTGAAGTTGGTTTGGAGCAAGCGGCACA 6653
DB 2310 TCTGAGAAACAGCGCCCTCTAATTTCACTTTGTGAAGTTGGTTTGGAGCAAGCGGCACA 2369
QY 6654 TGAACCAACCTATGATTTGTTTCACTGAGTGTGAGTTGGAAAGTCTGAGTTTAT 6713
DB 2370 TGAACCAACCTATGATTTGTTTCACTGAGTGTGAGTTGGAAAGTCTGAGTTTAT 2429
QY 6714 TGTCTGAGCACTTTTAAACCAACATATAATGACCAATGATTTTGTGATATATATGACT 6773
DB 2430 TGTCTGAGCACTTTTAAACCAACATATAATGACCAATGATTTTGTGATATATATGACT 2489
QY 6774 AGTAGCTGAACCTGAGAAAGTGAAGATGTCATGTCAGAAATCTGGCAGATATATCTT 6833
DB 2490 AGTAGCTGAACCTGAGAAAGTGAAGATGTCATGTCAGAAATCTGGCAGATATATCTT 2549
QY 6834 TTTTACACAGTGTCTCTGATCTCTTATCAATAAAGGAAAGTAAATCAGCCCATCTGTTT 6893
DB 2550 TTTTACACAGTGTCTCTGATCTCTTATCAATAAAGGAAAGTAAATCAGCCCATCTGTTT 2609
QY 6894 TGTTAACCTATTCAGCACTTCAGAAAGTGAAGTCTTTTGGAGCCCATGGAAGGTGATGTTGA 6953
DB 2610 TGTTAACCTATTCAGCACTTCAGAAAGTGAAGTCTTTTGGAGCCCATGGAAGGTGATGTTGA 2669
QY 6954 GCTTGAATGGGAGAAACCACTATGTAAATATTTTCAACCAAGGATATCAATTTGGAAGAGA 7013
DB 2670 GCTTGAATGGGAGAAACCACTATGTAAATATTTCAACCAAGGATATCAATTTGGAAGAGA 2729
QY 7014 TTTTAAATCCAGGGGCAAAAGTTAACCCCTCATTTCTCCGAAATTTGGAATGTCACCT 7073
DB 2730 TTTTAAATCCAGGGGCAAAAGTTAACCCCTCATTTCTCCGAAATTTGGAATGTCACCT 2789
QY 7074 TAAAGAAATATCTATGCTTCTCTCTCACTGTGCTTTT 7108
DB 2790 TAAAGAAATATCTATGCTTCTCTCTCACTGTGCTTTT 2824

RESULT 10
AX86451
LOCUS AX86451 2565 bp DNA linear PAT 29-MAR-2003
DEFINITION Sequence 7 from Patent WO02057450.
ACCESSION AX86451

VERSION	AX686451.1	GI:29372156	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1.		
AUTHORS	Edinger, S., Macdougall, J.R., Millet, I., Ellerman, K., Stone, D.J., Gerlach, V., Grose, W.M., Alsbrook, J.P., Lepley, D.M., Rieger, D., Burgess, C.E., Casman, S.J., Spytek, K.A., Boldog, F.L., Li, L., Padigara, M., Mishra, V., Patturajan, M., Shenoy, S., Rastelli, L., Tchernev, V.T., Vernet, C.A., Zerhusen, B.D., Malyankar, U.M., Guo, Y., Miller, C.E. and Gangolli, E.A.		
TITLE	Proteins and nucleic acids encoding same		
JOURNAL	Patent: WO 02057450-A 7 25-JUL-2002;		
FEATURES	Curagen Corporation (US)		
source	Location/Qualifiers		
	1..2565		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
	Query Match 35.9%; Score 2551; DB 6; Length 2565;		
	Best Local Similarity 99.8%; Pred. No. 0;		
	Matches 2554; Conservative 0; Mismatches 5; Indels 0; Gaps 0;		
QY	3262	CCTGAAGGGTTTCTTGGAAACCTGACTTACGAATCCATTTTGGTCAACCTGCAATATAATGTA	3321
DB	7	CCTGAAGGGTTTCTTGGAAACCTGACTTACGAATCCATTTTGGTCAACCTGCAATATAATGTA	66
QY	3322	AGTGGGTCCACCGGCTCAACCAAGCGTCTAGTCTTCTATGTTTCACTGATCTTA	3381
DB	67	AGTGGGTCCACCGGCTCAACCAAGCGTCTAGTCTTCTATGTTTCACTGATCTTA	126
QY	3382	CAGCAGACTCCTCGCCATGTGAGACCACTTCTGTTACATATGAGAGACATATATTT	3441
DB	127	CAGCAGACTCCTCGCCATGTGAGACCACTTCTGTTACATATGAGAGACATATATTT	186
QY	3442	GATAATCTGGAAAAATACACTGATATATATTAATAAATTAATCTCCATCAACAGAAAAAGGGA	3501
DB	187	GATAATCTGGAAAAATACACTGATATATATTAATAAATTAATCTCCATCAACAGAAAAAGGGA	246
QY	3502	TTCTCTGATACCTATATCTGCCAGCTATACATCAAGCTGAAGAAGATGCCAGAACT	3561
DB	247	TTCTCTGATACCTATATCTGCCAGCTATACATCAAGCTGAAGAAGATGCCAGAACT	306
QY	3562	TCACCAATATCAACTTTTAAACCTTTCTCTACCTCAGTCTCTTATCATGGAT	3621
DB	307	TCACCAATATCAACTTTTAAACCTTTCTCTACCTCAGTCTCTTATCATGGAT	366
QY	3622	CCCCAGTAAGCAAAATGGTGCATAATAAGTTATGATTTAACTTTTACAAGGACCAAAAT	3681
DB	367	CCCCAGTAAGCAAAATGGTGCATAATAAGTTATGATTTAACTTTTACAAGGACCAAAAT	426
QY	3682	GAATAATATTTCTTCTATTAATCTCTGATAATTAACATAATATGGAAGAGCTTTTCAACATTT	3741
DB	427	GAATAATATTTCTTCTATTAATCTCTGATAATTAACATAATATGGAAGAGCTTTTCAACATTT	486
QY	3742	ACATTATATAGCTTTTGTGTCGCGCAGACACTAGAAAAGGACTTGGTCTTCCAGTAT	3801
DB	487	ACATTATATAGCTTTTGTGTCGCGCAGACACTAGAAAAGGACTTGGTCTTCCAGTAT	546
QY	3802	CTTTTCTTTTACACAGATGAGTGCAGTGGCTTAGCACCTCCACAAAATTTGACTTTAATC	3861
DB	547	CTTTTCTTTTACACAGATGAGTGCAGTGGCTTAGCACCTCCACAAAATTTGACTTTAATC	606
QY	3862	AACTGACTTTCAGACTTTTGTATGGCTGAATGAGGCCCAAGTCTCTTCCAGGTGTAT	3921
DB	607	AACTGACTTTCAGACTTTTGTATGGCTGAATGAGGCCCAAGTCTCTTCCAGGTGTAT	666
QY	3922	GTTAAAGTATATAGTTTAAATTCATGAACATGAACATGAACATATATATATAAGAT	3981
DB			
DB	667	GTTAAAGTATATAGTTTAAATTCATGAACATGAACATGAACATATATATATAAGAT	726
QY	3982	ATATCAGGATTTAAATCTGAAGCAAACTTTGGATGGAACTGAACTGAGCACTACTCT	4041
DB	727	ATATCAGGATTTAAATCTGAAGCAAACTTTGGATGGAACTGAACTGAGCACTACTCT	786
QY	4042	ATCCGTGATCTCGGTTTCCACCAAGTGGAAATGGCAATCAATTTAGTATAGTATGATAA	4101
DB	787	ATCCGTGATCTCGGTTTCCACCAAGTGGAAATGGCAATCAATTTAGTATAGTATGATAA	846
QY	4102	TTCAACAACCAAGAATCAGTTCAGATGTCGTCGCAATATGCAAGTGCATGGCAACTAGC	4161
DB	847	TTCAACAACCAAGAATCAGTTCAGATGTCGTCGCAATATGCAAGTGCATGGCAACTAGC	906
QY	4162	TGCACTCAGTTTCTAGTGAATGGATCCACCAAAAAGCAATGGAATATAAATGACAG	4221
DB	907	TGCACTCAGTTTCTAGTGAATGGATCCACCAAAAAGCAATGGAATATAAATGACAG	966
QY	4222	TATATGGTAAACAGTTGAAAGGAATTTCAAAAAGTTTCTCCCAAGATCATGTACACT	4281
DB	967	TATATGGTAAACAGTTGAAAGGAATTTCAAAAAGTTTCTCCCAAGATCATGTACACT	1026
QY	4282	TTCAATAAGCTTCTTGGCAATACCTCATATGTTCTTTAAAGTAAAGAGCTTCAACTCAGCT	4341
DB	1027	TTCAATAAGCTTCTTGGCAATACCTCATATGTTCTTTAAAGTAAAGAGCTTCAACTCAGCT	1086
QY	4342	GGTGAAGGTGATGAAGAAGCACTGCCATGTGACACACTACCTGAAACAGTTCCTCAGTGT	4401
DB	1087	GGTGAAGGTGATGAAGAAGCACTGCCATGTGACACACTACCTGAAACAGTTCCTCAGTGT	1146
QY	4402	CCCAACAATATTTGCTTTTCTGATGTTCTGATGTTCACTAGTGCACATGATGATGATAAGA	4461
DB	1147	CCCAACAATATTTGCTTTTCTGATGTTCTGATGTTCACTAGTGCACATGATGATGATAAGA	1206
QY	4462	CCTGACACTATCTTGGCTACTTTTCAAAATTAACAAATTTACCACTCAACTCTGCTCTAT	4521
DB	1207	CCTGACACTATCTTGGCTACTTTTCAAAATTTACAAATTTACCACTCAACTCTGCTCTAT	1266
QY	4522	AAATGCAAAAGATGGGAATCCGAAGATGTTTGAATATCAAAAATTTCAATACCTCTAT	4581
DB	1267	AAATGCAAAAGATGGGAATCCGAAGATGTTTGAATATCAAAAATTTCAATACCTCTAT	1326
QY	4582	GAAGCTCACTTAATCGAAGAGACAGTATATGATTTAAAGAAATTTAGATGGTATGATTC	4641
DB	1327	GAAGCTCACTTAATCGAAGAGACAGTATATGATTTAAAGAAATTTAGATGGTATGATTC	1386
QY	4642	CAAGTGGCTGCCAGACCAATGCTGCTATGCAATGCTTCAAACTGGATTTCTACAAA	4701
DB	1387	CAAGTGGCTGCCAGACCAATGCTGCTATGCAATGCTTCAAACTGGATTTCTACAAA	1446
QY	4702	ACTCTGCTGGCTCCAGATGCTCTCTGAAAATGTTTCACTAGTAGCAACATCACCT	4761
DB	1447	ACTCTGCTGGCTCCAGATGCTCTCTGAAAATGTTTCACTAGTAGCAACATCACCT	1506
QY	4762	TTTAGCATCAGCATAGCTGGAGTGAACTGCTGTCATTTACTGGAACAACATGTTATCTG	4821
DB	1507	TTTAGCATCAGCATAGCTGGAGTGAACTGCTGTCATTTACTGGAACAACATGTTATCTG	1566
QY	4822	ATTGATGTCAAATCGGTAGATTAATGATGAAATTAATATATCTTCAAGTCAAAATGAA	4881
DB	1567	ATTGATGTCAAATCGGTAGATTAATGATGAAATTAATATATCTTCAAGTCAAAATGAA	1626
QY	4882	GAATAATAAACCATAGAAATTTAGAAATTTAGAAATTTTCAAGGATTTCTGATGATC	4941
DB	1627	GAATAATAAACCATAGAAATTTAGAAATTTAGAAATTTTCAAGGATTTCTGATGATC	1686
QY	4942	ACTGCATTTTCTGGGAACATTTAGTGTCTGATATGATGAGAGGGAAGTCAAGTGTGAAATG	5001
DB	1687	ACTGCATTTTCTGGGAACATTTAGTGTCTGATATGATGAGAGGGAAGTCAAGTGTGAAATG	1746
QY	5002	ATTGTTACTTCTTTAGAAATCAGCCCAAGGACCCCACTTAAACATGATCATTTTCAAG	5061
DB	1747	ATTGTTACTTCTTTAGAAATCAGCCCAAGGACCCCACTTAAACATGATCATTTTCAAG	1806


```
QY 5062 ATACCAGATGAGTTACAAAATTTCAATTAAGTTCTCTTCCCTCTCTCAACCTAATGCA 5121
Db 1807 ATACCAGATGAGTTACAAAATTTCAATTAAGTTCTCTTCCCTCTCTCAACCTAATGCA 1866

QY 5122 AATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCTTACTGCTGTCCAGATT 5181
Db 1867 AATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCTTACTGCTGTCCAGATT 1926

QY 5182 CACAACCTCAGTATTATACAGAAAACCAACATTCGTCTCAATTCGTAAGGACTA 5241
Db 1927 CACAACCTCAGTATTATACAGAAAACCAACATTCGTCTCAATTCGTAAGGACTA 1986

QY 5242 AAGGTGGACATACATCAATATACATGTTTACGCAAGTCAATAGTCTGTGGTCCAGTCCA 5301
Db 1987 AAGGTGGACATACATCAATATACATGTTTACGCAAGTCAATAGTCTGTGGTCCAGTCCA 2046

QY 5302 AAGGTTCCGATGAGATAACCATGGATATCAAAAGCTCCAGCAGACCAAAAACCAACCA 5361
Db 2047 AAGGTTCCGATGAGATAACCATGGATATCAAAAGCTCCAGCAGACCAAAAACCAACCA 2106

QY 5362 ACCCTTATTATGATGCCACAGAAAACCTGTTGTGACTTCAACAAATTAACAATCAGA 5421
Db 2107 ACCCTTATTATGATGCCACAGAAAACCTGTTGTGACTTCAACAAATTAACAATCAGA 2166

QY 5422 ATGCCAATATGTTACTACAGTATGATCATGACCAATTAATAAATGTACAAGTCTGCG 5481
Db 2167 ATGCCAATATGTTACTACAGTATGATCATGACCAATTAATAAATGTACAAGTCTGCG 2226

QY 5482 ACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGATGCATATTTAAT 5541
Db 2227 ACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGATGCATATTTAAT 2286

QY 5542 AAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAAACCTCCATGTACAGAGAAAG 5601
Db 2287 AAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAAACCTCCATGTACAGAGAAAG 2346

QY 5602 ACAAGTGTAGTGGCAATGAAGAAATCTACATAGTGTCTGATATATGCATCATGATT 5661
Db 2347 ACAAGTGTAGTGGCAATGAAGAAATCTACATAGTGTCTGATATATGCATCATGATT 2406

QY 5662 CTGGCAATGAACAAAAATTTTCAATGGACCACTGAAACCAAAAAAGCAATATTTT 5721
Db 2407 CTGGCAATGAACAAAAATTTTCAATGGACCACTGAAACCAAAAAAGCAATATTTT 2466

QY 5722 AAATTTAGAGCTACAAATATATGGGACAAATTTACTGACTGTGATTAATCTGACCTGTT 5781
Db 2467 AAATTTAGAGCTACAAATATATGGGACAAATTTACTGACTGTGATTAATCTGACCTGTT 2526

QY 5782 AAGACTTTAGGGAAGGACTTTTCAGAAAGACCGTAGAG 5820
Db 2527 AAGACTTTAGGGAAGGACTTTTCAGAAAGACCGTCGAG 2565

RESULT 11
LOCUS CQ726350
DEFINITION Sequence 12284 from Patent WO02068579.
ACCESSION CQ726350
VERSION CQ726350.1 GI:42289149
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 12284 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
```

```
source 1. .3149
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 31.7%; Score 2251; DB 6; Length 3149;
Best Local Similarity 80.5%; Pred. No. 0;
Matches 3020; Conservative 0; Mismatches 5; Indels 726; Gaps 4;

QY 2784 GAATAGATCATCATTAATAAACTATTAAATGTCATCTGAAACATCATTTGGAGATTATCAGATTT 2843
Db 1 GAATAGATCATCATTAATAAACTATTAAATGTCATCTGAAACATCATTTGGAGATTATCAGATTT 60

QY 2844 GGAATTATATGTTGAATACAGTCTTATGTACAGCTAGCACCAAGATTTGGTATGGGAA 2903
Db 61 GGAATTATATGTTGAATACAGTCTTATGTAAACAGCTAGCACCAAGATTTGGTATGGGAA 120

QY 2904 AACAGAGCAATATATCATTAGCTTTCAAAACACAGAGGAGGACCAAGCGATCTCTCCAA 2963
Db 121 AACAGAGCAATATATCATTAGCTTTCAAAACACAGAGGAGGACCAAGCGATCTCTCCAA 180

QY 2964 AGATGTTTATTATGCAAACTCAGTTCCTCATCAATAATTTCTTTCTGGACACCTCTTC 3023
Db 181 AGATGTTTATTATGCAAACTCAGTTCCTCATCAATAATTTCTTTCTGGACACCTCTTC 240

QY 3024 AAAACCTAATGGATATACATATTAATCTCTGTTTATTACAGAAATACCTTCAGGTACTTT 3083
Db 241 AAAACCTAATGGATATACATATTAATCTCTGTTTATTACAGAAATACCTTCAGGTACTTT 300

QY 3084 TATGCAAGATTTTACACTCCATGAACCTAATGACTTTGACAAATATGACTGTATCCAC 3143
Db 301 TATGCAAGATTTTACACTCCATGAAGTAAACCAATGACTTTGACAAATATGACTGTATCCAC 360

QY 3144 AATTATAGATAAATGCAATATTTACAGCTAATATACATTTTGGTTAAACAGAGTACTTC 3203
Db 361 AATTATAGATAAATGCAATATTTACAGCTAATATACATTTTGGTTAAACAGAGTACTTC 420

QY 3204 AGTTCGAATGGGATTAAGCAGTACATCATTTGAAGTATACACAGATCAAGACATACC 3263
Db 421 AGTTCGAATGGGATTAAGCAGTACATCATTTGAAGTATACACAGATCAAGACATACC 476

QY 3264 TGAAGGTTTGTGGAAACCTGACTTACGAATCCATTTCTGTCCTCAACTGAATTAAG 3323
Db 477 ----- 476

QY 3324 CTGGTCCACCCGGCTCAACAAAGGTCTAGTCTTCTACTATGTTTCACTGATCTTACA 3383
Db 477 ----- 476

QY 3384 GCAGACTCTCGCCATGTGAGACCCTCTTGTGTATATGAGAGAAGCATATATTTTGA 3443
Db 477 ----- 476

QY 3444 TAATCTGGAAAAATAACACTGATTATATATAATAAATACTCCATCAACAGAAAAAGGATT 3503
Db 477 ----- 476

QY 3504 CTCTGATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAGATGTCCAGAAACTTC 3563
Db 477 ----- 489

QY 3564 ACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTTCCTTTATCATGGGATCC 3623
Db 490 ACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTTCCTTTATCATGGGATCC 549

QY 3624 CCAGTAAAGCCAAATGGTGGCAATAATAAGTTATGATTTAACTTTAAAGGACCAATGA 3683
Db 550 CCAGTAAAGCCAAATGGTGGCAATAATAAGTTATGATTTAACTTTAAAGGACCAATGA 609

QY 3684 AAATTTCTTTTCATTTACTCTTGATTAATTTACATAATTTGGAAGAGCTTTCCACATTTAC 3743
Db 610 AAATTTCTTTTCATTTACTCTTGATTAATTTACATAATTTGGAAGAGCTTTCCACATTTAC 669
```

QY 3744 ATATATAGCTTTTCTGCTCCGCAAGAACTAGAAAAGGACTTGGTCTCTTCCAGTATCT 3803
DB 670 ATATATAGCTTTTCTGCTCCGCAAGAACTAGAAAAGGACTTGGTCTCTTCCAGTATCT 729
QY 3804 TTTCTTTTACACAGATGAGTCAGTGGCGTTAGCACCTCCACAAAATTTGACTTTTAATCAA 3863
DB 730 TTTCTTTTACACAGATGAGTCAGTGGCGTTAGCACCTCCACAAAATTTGACTTTTAATCAA 789
QY 3864 CTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGTGGTATGT 3923
DB 790 CTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGTGGTATGT 849
QY 3924 TAAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATATATATATGAATAT 3983
DB 850 TAAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATATATATATGAATAT 909
QY 3984 ATCAGGATTTAAACTGAAGCCAACTTTGTTGAGCTGGAACCACTGACGACCTACTCTAT 4043
DB 910 ATCAGGATTTAAACTGAAGCCAACTTTGTTGAGCTGGAACCACTGACGACCTACTCTAT 969
QY 4044 CCGTGATCTGCGTTACCAAAAGTTGAAATGGCAATCAATTTAGTAATGTAGTAAAT 4103
DB 970 CCGTGATCTGCGTTACCAAAAGTTGAAATGGCAATCAATTTAGTAATGTAGTAAAT 1029
QY 4104 CACAAACCAAGATCAGTTCAGATGTCGTCAGAAATATGAGTGCATGCAACTAGCTG 4163
DB 1030 CACAAACCAAGATCAGTTCAGATGTCGTCAGAAATATGAGTGCATGCAACTAGCTG 1089
QY 4164 GCAGTCAAGTTTGTAGTGAATGGGATCCACCCAAAAGGCAATGGATATATACGCGAT 4223
DB 1090 GCAGTCAAGTTTGTAGTGAATGGGATCCACCCAAAAGGCAATGGATATATACGCGAT 1149
QY 4224 TATGGTAAACAGTTGAAAGGAATCTACAAAAGTTTCTCCCAAGATCACATGTACACTTT 4283
DB 1150 TATGGTAAACAGTTGAAAGGAATCTACAAAAGTTTCTCCCAAGATCACATGTACACTTT 1209
QY 4284 CATAAAGCTTTCTGCAATACCTCATATGTTCTTTAAAGTAAGAGCTTCAACCTCAGCTG 4343
DB 1210 CATAAAGCTTTCTGCAATACCTCATATGTTCTTTAAAGTAAGAGCTTCAACCTCAGCTG 1269
QY 4344 TGAAGTGTAAGAACACATGCCATGTGACACACATCTACCTGAAACAGTTCACAGTGTCC 4403
DB 1270 TGAAGTGTAAGAACACATGCCATGTGACACACATCTACCTGAAACAG- 1316
QY 4404 CACAAATATTGCTTTCTGATGTTCACTGATCACTAGTGCAACATTGACATGGATAAGACC 4463
DB 1317 ----- 1316
QY 4464 TGACACTATCTTTGGCTACTTTCAAAATTAACAAATTAACACTTCTGCTCAAAA 4523
DB 1317 ----- 1316
QY 4524 ATGCAAGATGGGAATCCGAAGATGTGTTGAATATCAAAAATTCATACCTCTATGA 4583
DB 1317 ----- 1316
QY 4584 AGCTCACTTAACCTGAAGAGACAGTATATGATTAAGAAATTTAGATGGTATAGATTCCA 4643
DB 1317 ----- 1316
QY 4644 AGTGGCTGCAGACCAATGCTGGCTATGSCAATGCTTCAAACTGGATTTCTACAAAAAC 4703
DB 1317 ----- 1316
QY 4704 TCTGCTGGCCCTCCAGATGGTCTCTCTGAAATGTTTATGATAGTAGCAACATCAGCTTT 4763
DB 1317 -----CTCCAGATGGTCTCTCTGAAATGTTTATGATAGTAGCAACATCAGCTTT 1365
QY 4764 TAGCATCAGCATAGCTGGAGTCAACTGCTGTCATTTACTGGACCAACATGTTATCTGAT 4823
DB 1366 TAGCATCAGCATAGCTGGAGTCAACTGCTGTCATTTACTGGACCAACATGTTATCTGAT 1425

QY 4824 TGATGTCAAAATCGGTAGATAATGATGAATTTAATATATATCTTTCATCAAGTCAAAATGAAGA 4883
DB 1426 TGATGTCAAAATCGGTAGATAATGATGAATTTAATATATATCTTTCATCAAGTCAAAATGAAGA 1485
QY 4884 AATAAAAAACCATAGAAATTAAGATTTAGAAATATTTACAAAGGTATTTCTGTAGTGATCAC 4943
DB 1486 AATAAAAAACCATAGAAATTAAGATTTAGAAATATTTCAAAAGGTATTTCTGTAGTGATCAC 1545
QY 4944 TGCATTTACTGGGAACATTTAGTGTCTCATATGTAGAAAGGGAAGTCAAGTGTCTGAAATGAT 5003
DB 1546 TGCATTTACTGGGAACATTTAGTGTCTCATATGTAGAAAGGGAAGTCAAGTGTCTGAAATGAT 1605
QY 5004 TGTACTACTTTTAGAATCAGCCCCCAAGGACCCACTAACACATGACATTTTTCAGAAATGAT 5063
DB 1606 TGTACTACTTTTAGAATC----- 1623
QY 5064 ACCAGATGAAGTTTACAAAATTTCAAATTAAGCTTCTCTTCTCTCAACCTAATGGAAA 5123
DB 1624 ----- 1623
QY 5124 TATCCAAGTATATCAAAGCTCTGGTTTACCGAGAAGATGATCTTCTGCTGTCCAGATTTCA 5183
DB 1624 -----AGTATATCAAAGCTCTGGTTTACCGAGAAGATGATCTTCTGCTGTCCAGATTTCA 1677
QY 5184 CAACCTCAGTATTTATACAGAAAACCAACACATCTGCTGATTCGAAATGCTAGAGGACTAAA 5243
DB 1678 CAACCTCAGTATTTATACAGAAAACCAACACATCTGCTGATTCGAAATGCTAGAGGACTAAA 1737
QY 5244 AGGTGACACATACATACATATCAGTGTCTTACGCAAGTCAATAGTGTGCTGCTGCTCAAA 5303
DB 1738 AGGTGACACATACATACATATCAGTGTCTTACGCAAGTCAATAGTGTGCTGCTGCTCNA 1797
QY 5304 GGTTCCGATGAGAATTAACCATGGATATCAAAGCTCCAGACGACCAAAAACCAAAACCAAC 5363
DB 1798 GGTTCCGATGAGAATTAACCATGGATATCAAAGCTCCAGACGACCAAAAACCAAAACCAAC 1857
QY 5364 CCTATTTATGATGCCACAGGAAAACCTGCTTGTGACTTCAAACATTTACAATTCAGAT 5423
DB 1858 CCTATTTATGATGCCACAGGAAAACCTGCTTGTGACTTCAAACATTTACAATTCAGAT 1917
QY 5424 GCCAATATCTTACTACAGTGTATCATGACCAATAAATAAATGTAACAGTGTCTGCGAC 5483
DB 1918 GCCAATATGTTACTACAGTGTATCATGACCAATAAATAAATGTAACAGTGTCTGCGAC 1977
QY 5484 AGAAAACAGGAGCTCAGCATGTGGAATGTAAACAAAGTGGTATGATGCTATTTTAAATA 5543
DB 1978 AGAAAACAGGAGCTCAGCATGTGGAATGTAAACAAAGTGGTATGATGCTATTTTAAATA 2037
QY 5544 AGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCTCCATGTACAGAGGAAAGAC 5603
DB 2038 AGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCTCCATGTACAGAGGAAAGAC 2097
QY 5604 ABAAGTTTGTGCGCAATGAAGAAATCTACATCATAGTGTGATTAATGCAATGCAATGATTC 5663
DB 2098 ABAAGTTTGTGCGCAATGAAGAAATCTACATCATAGTGTGATTAATGCAATGCAATGATTC 2157
QY 5664 TGGCAATGAAGCAAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATACCTTTTAA 5723
DB 2158 TGGCAATGAAGCAAAAATTTGCAATGGACCACTGAAACCAAAAAGCAATACCTTTTAA 2217
QY 5724 ATTTAGAGCTACAAATATTTATGGGCAATTTTACTGATCTGATTTTCTGACCCCTGTAA 5783
DB 2218 ATTTAGAGCTACAAATATTTATGGGCAATTTTACTGATCTGATTTTCTGACCCCTGTAA 2277
QY 5784 GACTTTAGGGGAGGACTTTTACAGAAAGACCTGAGATCATTTCTTCCGTCACTTTTG 5843
DB 2278 GACTTTAGGGGAGGACTTTTACAGAAAGACCTGAGATCATTTCTTCCGTCACTTTTG 2337
QY 5844 TATCTTTTCAATAATTTCTCTTGGAAACAGCTATTTTGGCAATTTGCAAGAAATTCGACAGAA 5903
DB 2338 TATCTTTTCAATAATTTCTCTTGGAAACAGCTATTTTGGCAATTTGCAAGAAATTCGACAGAA 2397
QY 5904 GCAGAAAGAGGTGGACATATCTCTCTCAGGATGCAGAAATTTATTTGACACTAAATTTGAA 5963

Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;									
QY	6015	AAGCGGCGCAATAAGCAAGAAATCCCTCTCGAACAATGTTGAAGACTTTGCACAAACA	6074						
Db	70	AATGAGCCCAATAAGCAAGAAATCCCTCTCGAACAATGTTGAAGACTTTGCACAAACA	129						
QY	6075	CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC	6134						
Db	130	CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC	189						
QY	6135	AACTGATGCTGATCTGCCCTTGGAAATAGACAAACCGTTTCCCAAAATATAAAACCATP	6194						
Db	190	AACTGATGCTGATCTGCCCTTGGAAATAGACAAACCGTTTCCCAAAATATAAAACCATP	248						
QY	6195	TAATAATAATACAGAGTAAGCTGATAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT	6254						
Db	249	--ATAATAATAACAGAGTAAGCTGATAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT	306						
QY	6255	TAATGCCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC	6314						
Db	307	TAATGCCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC	366						
QY	6315	ACTACAGGAAAGTTTGGAGATTTTGGAGATTTGTTGGGAAACAGGCGCAAAACATP	6374						
Db	367	ACTACAGGAAAGTTTGGAGATTTTGGAGATTTGTTGGGAAACAGGCGCAAAACATP	426						
QY	6375	AGTAATGCTTAACACAGTGTTTGAAAAGGACGGATCAGATGCCATTCAGTATTTGGCCAGA	6434						
Db	427	AGTAATGCTTAACACAGTGTTTGAAAAGGACGGATCAGATGCCATTCAGTATTTGGCCAGA	486						
QY	6435	GGACAAACAGCCAGTTACTGCTTTGGAGATATAGTGATTAACAAGCTAATGGAGGATGT	6494						
Db	487	GGACAAACAGCCAGTTACTGCTTTGGAGATATAGTGATTAACAAGCTAATGGAGGATGT	546						
QY	6495	TCAAATAGATTGGACTATCAGGGATCTGAAAATTTGGAAGCGATGGGATTTGCATGACTGT	6554						
Db	547	TCAAATAGATTGGACTATCAGGGATCTGAAAATTTGGAAGCGATGGGATTTGCATGACTGT	606						
QY	6555	TCGACAGTGAATCTTCTGCGCCAGAGCATGGGTTCTGAGAAACAGCGCCCTCT	6614						
Db	607	TCGACAGTGAATCTTCTGCGCCAGAGCATGGGTTCTGAGAAACAGCGCCCTCT	666						
QY	6615	AATTCACCTTTGTAAGTTTGGAGAACTGGAGTTTTTATTTGCTCTGGACCAATTAACACA	6674						
Db	667	AATTCACCTTTGTAAGTTTGGAGAACTGGAGTTTTTATTTGCTCTGGACCAATTAACACA	726						
QY	6675	TCATCTGAGTCTGGAGTTGGAGAACTGGAGTTTTTATTTGCTCTGGACCAATTAACACA	6734						
Db	727	TCATCTGAGTCTGGAGTTGGAGAACTGGAGTTTTTATTTGCTCTGGACCAATTAACACA	786						
QY	6735	ACATAAATGACCATGATTTTGTGGATATATAGCTAGTATGCTAGTGAAGTGA	6794						
Db	787	ACATAAATGACCATGATTTTGTGGATATATAGCTAGTATGCTAGTGAAGTGA	846						
QY	6795	AGAATGTCATGGTGCAGATCTCGCACAGTATATCTTTTACACAGTCAATCTGGA	6854						
Db	847	AGAATGTCATGGTGCAGATCTCGCACAGTATATCTTTTACACAGTCAATCTGGA	906						
QY	6855	TCTCTTATCAAAATAAGGAGTAATCAGCCCATCTGTTTGTAACTATTTCAGCACTTCA	6914						
Db	907	TCTCTTATCAAAATAAGGAGTAATCAGCCCATCTGTTTGTAACTATTTCAGCACTTCA	966						
QY	6915	GAAGATGGACTCTTTTGGACGCCATGGAAGTGATGTTGAGCTTGAATGGGAGAAACCA	6974						
Db	967	GAAGATGGACTCTTTTGGACGCCATGGAAGTGATGTTGAGCTTGAATGGGAGAAACCA	1026						
QY	6975	TATGTAATATTCAGACCAAGAGATACAAATGGAACAGATTTTAAATCCAGGGGCCAA	7034						
Db	1027	TATGTAATATTCAGACCAAGAGATACAAATGGAACAGATTTTAAATCCAGGGGCCAA	1086						
QY	7035	AGTTACCCCTCAATCTTCCGAATTTGAAATGTCGAACCTTTAAGAAATATCTATGCTTCT	7094						
Db	1087	AGTTACCCCTCAATCTTCCGAATTTGAAATGTCGAACCTTTAAGAAATATCTATGCTTCT	1146						

Qy	7095	CTCACTGTGCGCTTT	7108						
Db	1147	CTCACTGTGCGCTTT	1160						
RESULT 15									
I32037	I32037	Sequence 10 from patent US 5585233.	2309 bp	DNA	linear	PAT 06-FEB-1997			
LOCUS	I32037	Sequence 10 from patent US 5585233.							
DEFINITION	I32037	Sequence 10 from patent US 5585233.							
ACCESSION	I32037	Sequence 10 from patent US 5585233.							
VERSION	I32037.1	GI:1822828							
KEYWORDS									
SOURCE	Unknown.								
ORGANISM	Unknown.								
REFERENCE	1 (bases 1 to 2309)								
AUTHORS	Moller,N.P.H., Moller,K.B. and Ullrich,A.								
TITLE	PTP-S31: A novel protein tyrosine phosphatase								
JOURNAL	Patent: US 5585233-A 10 17-DEC-1996;								
FEATURES	Location/Qualifiers								
source	1..2309								
	/organism="unknown"								
ORIGIN	/mol_type="unassigned DNA"								
Query Match 15.1%; Score 1071.6; DB 6; Length 2309;									
Best Local Similarity 99.4%; Pred. No. 4.7e-207;									
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;									
Qy	6015	AACGGCGCAATAAGCAAGAAATCCCTCTCGAACAATGTTGAAGAGCTTTGCACAAACA	6074						
Db	70	AATGAGCCCAATAAGCAAGAAATCCCTCTCGAACAATGTTGAAGAGCTTTGCACAAACA	129						
Qy	6075	CAACCTAAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC	6134						
Db	130	CAACCTAAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC	189						
Qy	6135	AACTGATGCTGATCTGCCCTTGGAAATAGACAAACCGTTTCCCAAAATATAAAACCATP	6194						
Db	190	AACTGATGCTGATCTGCCCTTGGAAATAGACAAACCGTTTCCCAAAATATAAAACCATP	248						
Qy	6195	TAATAATAATACAGAGTAAAGCTGATAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT	6254						
Db	249	--ATAATAATACAGAGTAAAGCTGATAGCTAGCTAGCTAGCTAGCTAGCTAGCT	306						
Qy	6255	TAATGCCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC	6314						
Db	307	TAATGCCAGCTATATTTCTGGTTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC	366						
Qy	6315	ACTACAGGAAAGTTTGGAGATTTTGGAGATTTGTTGGGAAACAGGCGCAAAACATP	6374						
Db	367	ACTACAGGAAACAGTTTGGAGATTTTGGAGATTTGTTGGGAAACAGGCGCAAAACATP	426						
Qy	6375	AGTAATGCTTAACACAGTGTTTGAAAAGGACGGATCAGATGCCATTCAGTATTTGGCCAGA	6434						
Db	427	AGTAATGCTTAACACAGTGTTTGAAAAGGACGGATCAGATGCCATTCAGTATTTGGCCAGA	486						
Qy	6435	GGACAAACAGCCAGTTACTGCTTTTGGAGATATAGTGATTAACAAGCTAATGGAGGATGT	6494						
Db	487	GGACAAACAGCCAGTTACTGCTTTTGGAGATATAGTGATTAACAAGCTAATGGAGGATGT	546						
Qy	6495	TCAAATAGATTGACCTATCAGGGATCTGAAAATTTGAAAGCGCATGGGATTTGCATGACTGT	6554						
Db	547	TCAAATAGATTGACCTATCAGGGATCTGAAAATTTGAAAGCGCATGGGATTTGCATGACTGT	606						
Qy	6555	TCGACAGTGAATCTTTTACTGCTCGCCAGAGCATGGGGTTCTTGAGAAACAGCGCCCTCT	6614						
Db	607	TCGACAGTGAATCTTTTACTGCTCGCCAGAGCATGGGGTTCTTGAGAAACAGCGCCCTCT	666						
Qy	6615	AATTCATTTTGAAGTTTGGTTTCAGCAAGCAGGCGCATGACACCCACACTATGATTGT	6674						
Db	667	AATTCATTTTGAAGTTTGGTTTCAGCAAGCAGGCGCATGACACCCACTATGATTGT	726						


```
Qy 6675 TCACTGCAGTGTGGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCAATTTAACACA 6734
Db |||||||
Qy 727 TCACTGCAGTGTGGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCAATTTAACACA 786
Db |||||||
Qy 6735 ACATATAAATGACCATGATTTCTGGATATATATGACTAGTAGTCTGAACTGGAAGTGA 6794
Db |||||||
Qy 787 ACATATAAATGACCATGATTTCTGGATATATATGACTAGTAGTCTGAACTGGAAGTGA 846
Db |||||||
Qy 6795 AAGAATGTGCATGGTGCAGAACTCTGGCAGTATATCTTTTACACCACTGCAATCTCTGGA 6854
Db |||||||
Qy 847 AAGATGTGCATGGTGCAGAACTCTGGCAGTATATCTTTTACACCACTGCAATCTCTGGA 906
Db |||||||
Qy 6855 TCTCTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTTCAGCACTTCA 6914
Db |||||||
Qy 907 TCTCTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTTCAGCACTTCA 966
Db |||||||
Qy 6915 GAAGATGGACTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAACCCAC 6974
Db |||||||
Qy 967 GAAGATGGACTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAACCCAC 1026
Db |||||||
Qy 6975 TATGTAATAATTTCAGACCAAGGATACAAATTGGAAGAGATTTTAAATCCAGGGGCCAA 7034
Db |||||||
Qy 1027 TATGTAATAATTTCAGACCAAGGATACAAATTGGAAGAGATTTTAAATCCAGGGGCCAA 1086
Db |||||||
Qy 7035 AGTTACCCCTCATTTCTTCGAAATTGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT 7094
Db |||||||
Qy 1087 AGTTACCCCTCATTTCTTCGAAATTGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT 1146
Db |||||||
Qy 7095 CTCACGTGCGCTTT 7108
Db |||||||
Qy 1147 CTCACGTGCGCTTT 1160
Db |||||||
```

Search completed: July 2, 2005, 12:07:40
Job time : 48096 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 19:01:59 ; Search time 5203 Seconds
(without alignments)
8087.165 Million cell updates/sec

Title: US-10-673-885-1
Perfect score: 7108
Sequence: 1 taattgtactgtccagaa.....gcttctcactgtgccttt 7108

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7108	100.0	7108	8 ABZ23703	Abz23703 Human pho
2	6800.8	95.7	6897	9 AAD56411	Aad56411 Human rec
3	6799.8	95.7	7225	8 ABA00717	Abao0717 Human KPP
4	6799	95.7	6900	12 ADH41614	Adh41614 Novel hum
5	6683.6	94.0	6903	6 ABT06283	Abt06283 Human NOV
6	6683.6	94.0	6903	12 ADH41624	Adh41624 Novel hum
7	6453	90.8	6994	6 ABT06281	Abt06281 Human NOV
8	6453	90.8	6994	12 ADH41618	Adh41618 Novel hum
9	6340	89.2	6881	12 ADH41612	Adh41612 Novel hum
10	5619.4	79.1	5877	9 AAD56413	Aad56413 Human rec
11	2716.2	38.2	2739	12 ADH41620	Adh41620 Novel hum
12	2551	35.9	2565	6 ABT06282	Abt06282 Human NOV
13	2551	35.9	2565	12 ADH41622	Adh41622 Novel hum
14	2205	31.0	2220	12 ADH41616	Adh41616 Novel hum
15	1071.6	15.1	2309	2 AAQ73786	Aaq73786 Partial p
16	905.2	12.7	990	9 AAD56412	Aad56412 Human rec
17	532	7.5	532	2 AAQ73785	Aaq73785 Partial p
18	493.6	6.9	1380	12 ADP29100	Adp29100 Human sec
19	319.4	4.5	321	2 AAQ73782	Aaq73782 Partial p
20	315	4.4	254366	8 ABZ23704	Abz23704 Human pho

C	21	230.8	3.2	412	9	AAD56414	Aad56414 Human rec
	22	195.2	2.7	3623	2	AAV82821	Aav82821 Rabbit os
	23	195.2	2.7	5679	2	AAV04867	Aav04867 Nucleotid
	24	193.8	2.7	6329	9	ACC79776	Acc79776 Human PTP
	25	192.2	2.7	936	12	ADO04583	Ado04583 Human HPT
	26	192.2	2.7	1011	12	ADO04581	Ado04581 Human HPT
	27	192.2	2.7	3984	12	ADH48408	Adh48408 Human KPP
	28	192.2	2.7	4752	2	AAV04868	Aav04868 Nucleotid
	29	192.2	2.7	4815	2	AAV04886	Aav04886 Nucleotid
	30	192.2	2.7	5069	10	ADD89070	Add89070 Encoding
	31	192.2	2.7	5113	2	AAT48479	Aat48479 Human tyr
	32	192.2	2.7	5153	13	ADP54714	Adp54714 Human pro
	33	192.2	2.7	5153	13	ADP24083	Adp24083 PRO polypp
	34	192.2	2.7	5415	13	ADK52802	Adk52802 Drug ther
	35	192.2	2.7	5415	13	ADS17077	Ads17077 Human 207
	36	192.2	2.7	6075	3	AAA88866	Aaa88866 Human pro
	37	192.2	2.7	6075	11	ADI32040	Adi32040 Human cDN
	38	192.2	2.7	6075	12	ADO04579	Ado04579 Human HPT
	39	192.2	2.7	6342	4	AAK51954	Aak51954 Human pol
	40	192.2	2.7	6632	4	AAL26742	Aal26742 Human bre
	41	192.2	2.7	7444	6	AAD32364	Aad32364 Human lun
	42	192.2	2.7	7444	10	ADE54130	Ad54130 Human pro
	43	192.2	2.7	7879	11	ACN88701	Acn88701 Breast ca
	44	187.4	2.6	4799	2	AAV04887	Aav04887 Nucleotid
	45	184.4	2.6	6346	4	AAK52938	Aak52938 Human pol

ALIGNMENTS

RESULT 1
ABZ23703
ID ABZ23703 standard; cDNA; 7108 BP.

XX ABZ23703;

DT 21-MAR-2003 (first entry)

DE Human phosphatase encoding cDNA #SEQ ID 1.

KW Human; phosphatase; enzyme; chromosome 12; colon adenocarcinoma;
KW placenta; gene therapy; gene; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT 5'UTR 1..105

FT CDS /*tag= a

FT /*tag= b

FT /*product= "phosphatase protein"

FT 3'UTR 6982..7108

FT /*tag= C

XX WO200279452-A2.

XX 10-OCT-2002.

XX 02-APR-2002; 2002WO-US009992.

XX 02-APR-2001; 2001US-00822871.

XX (PEKE) PE CORP NY.

XX Webster M, Wei M, Di Francesco V, Beasley EM;

XX WPI; 2003-046809/04.

XX P-PSDB; ABP60057.

PT New human phosphatase peptides and nucleic acids encoding the peptides,
PT useful as models in developing human therapeutic targets, in identifying
PT therapeutic proteins or modulators of protein activity, and in gene
PT therapy.

XX PS Claim 4b; Fig 1; 220bp; English.

XX CC The invention relates to an isolated human phosphatase peptide. The human phosphatase peptides and nucleic acids are useful as models for the development of human therapeutic targets and in the identification of therapeutic proteins. They also serve as targets for the development of human therapeutic agents that modulate phosphatase activity in cells and tissues that express the phosphatase. Experimental data indicates that the cDNA is expressed in colon adenocarcinoma and placenta. The proteins can be used to raise or to elicit another immune response, as a reagent in assays designed to determine the levels of the protein in biological fluids, as markers for tissue in which the corresponding protein is preferentially expressed, in the identification of modulators of the peptides, and in pharmacogenomic analysis. The polypeptides and polynucleotides may be used in gene therapy, and as antisense constructs to control phosphatase gene expression in cells, tissue and organisms. The gene encoding the phosphatase of the invention is located on human chromosome 12. The current sequence represents the human phosphatase encoding cDNA

XX SQ Sequence 7108 BP; 2387 A; 1408 C; 1334 G; 1979 T; 0 U; 0 Other;

Query Match 100.0%; Score 7108; DB 8; Length 7108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAATTGTACTGCGCAGGAGTCTGCTTTAAATCAATTAATGACGGCAACATTTCTCT 60
DB 1 TAATTGTACTGCGCAGGAGTCTGCTTTAAATCAATTAATGACGGCAACATTTCTCT 60

QY 61 CTAGAGCCATCATGTGATCTACTGCTGAAATGTAATGAAGTGAATTTCTTATC 120
DB 61 CTAGAGCCATCAATGATGATCTACTGCTGAAATGTAATGAAGTGAATTTCTTATC 120

QY 121 AATTTTCTTTTACTTTTATGCGACTTCAGACACACAGGTGATGTTCCAAATGTCGT 180
DB 121 AATTTTCTTTTACTTTTATGCGACTTCAGACACACAGGTGATGTTCCAAATGTCGT 180

QY 181 CTTGGTACTAGGTACGATATAACATCTCTTTCAATTTCTACAAATACACCTCCTGTT 240
DB 181 CTTGGTACTAGGTACGATATAACATCTCTTTCAATTTCTACAAATACACCTCCTGTT 240

QY 241 ACTAGAAATAGTACACCAATGTAACAAACACAGGCGCTCCAGTCTTCTAGCCGGGAA 300
DB 241 ACTAGAAATAGTACACCAATGTAACAAACACAGGCGCTCCAGTCTTCTAGCCGGGAA 300

QY 301 AGAGTCGGATCTGCTGGGATTTCTGCTTTGGAATACACCACTTAATCCAAATGGAAG 360
DB 301 AGAGTCGGATCTGCTGGGATTTCTGCTTTGGAATACACCACTTAATCCAAATGGAAG 360

QY 361 ATTATATCTTACATTTGTCMAATTAAGGAATTTGTCGGTGGATGCAACAGTATATACA 420
DB 361 ATTATATCTTACATTTGTCMAATTAAGGAATTTGTCGGTGGATGCAACAGTATATACA 420

QY 421 CRAGTGAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAATCTTAATCTCGGAACA 480
DB 421 CRAGTGAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTACTAATCTTAATCTCGGAACA 480

QY 481 ACATATGAAATTAAGGTTGCTGCTGAAAAAGTGTGCAATTTGAGTGTGATGATCCA 540
DB 481 ACATATGAAATTAAGGTTGCTGCTGAAAAAGTGTGCAATTTGAGTGTGATGATCCA 540

QY 541 TTTTCTCTCAAACTGCGAGAAAGTCTCCAGGAAAAAGTGTGCAATTTGAGTGTGATGATCCA 600
DB 541 TTTTCTCTCTCAAACTGCGAGAAAGTCTCCAGGAAAAAGTGTGCAATTTGAGTGTGATGATCCA 600

QY 601 TACAACGCTTCAGCAGTTAGCTGATTTGTTATTTACCTCGGCAACCAATGCAAAAT 660
DB 601 TACAACGCTTCAGCAGTTAGCTGATTTGTTATTTACCTCGGCAACCAATGCAAAAT 660

QY 661 ACCAGCTTCAAGATTTAGTGTCAAAACATGCCAGAAAGTGGATAGTGTGAAAGATGTTCTCA 720
DB 661 ACCAGCTTCAAGATTTAGTGTCAAAACATGCCAGAAAGTGGATAGTGTGAAAGATGTTCTCA 720

DB 661 ACCAGCTTCAAGATTTAGTGTCAAAACATGCCAGAAAGTGGATAGTGTGAAAGATGTTCTCA 720

QY 721 ATCAGAGTAGAGGACATTTTGAATCTGGGAAATTTGCCAGAAATGCAATGAGATAGTGAATCT 780
DB 721 ATCAGAGTAGAGGACATTTTGAATCTGGGAAATTTGCCAGAAATGCAATGAGATAGTGAATCT 780

QY 781 TTTTATGAGTAGACAGCAGCCCTTCTCAACCCCTTGGTAGAGTTACACCTCCTCATCGCT 840
DB 781 TTTTATGAGTAGACAGCAGCCCTTCTCAACCCCTTGGTAGAGTTACACCTCCTCATCGCT 840

QY 841 ACCACACATTTCAAGCAGCTTGACACAGATAGATACAGCTCTGTGTGGAAGAGCT 900
DB 841 ACCACACATTTCAAGCAGCTTGACACAGATAGATACAGCTCTGTGTGGAAGAGCT 900

QY 901 ATCAGTTTGTAGTACACACCTTGAGACCTTATACACATATCTTTTGAAGTTTCAGCT 960
DB 901 ATCAGTTTGTAGTACACACCTTGAGACCTTATACACATATCTTTTGAAGTTTCAGCT 960

QY 961 GCTACAACTGAAGCAGGTTTATATGATAGTACGATTTGTGAGAACTCAGATGCT 1020
DB 961 GCTACAACTGAAGCAGGTTTATATGATAGTACGATTTGTGAGAACTCAGATGCT 1020

QY 1021 GAAGGACCAACCAAACTGCGTAAACAGCAACATCAAGGAAAGTCTTTTCAATTTTA 1080
DB 1021 GAAGGACCAACCAAACTGCGTAAACAGCAACATCAAGGAAAGTCTTTTCAATTTTA 1080

QY 1081 TGGGACCCACCACTTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTTATGAGCA 1140
DB 1081 TGGGACCCACCACTTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTTATGAGCA 1140

QY 1141 TCAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAGTTTGCATTCTCAATCACTAA 1200
DB 1141 TCAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAGTTTGCATTCTCAATCACTAA 1200

QY 1201 CCATTTACAATGATGATGTTCTATATTCGGCTGAAAACAGTGCAGGAGCTGGGCCAAG 1260
DB 1201 CCATTTACAATGATGATGTTCTATATTCGGCTGAAAACAGTGCAGGAGCTGGGCCAAG 1260

QY 1261 TCAATATTTTCAATGATGTTCTATATTCAGGAGTGTTCAGGGGAGTGTGATTTTCAACT 1320
DB 1261 TCAATATTTTCAATGATGTTCTATATTCAGGAGTGTTCAGGGGAGTGTGATTTTCAACT 1320

QY 1321 GCAGAGGTAGAAATCCACGCAAGTAAAGAAATTTCTTGGAAAGAAACACGCAACCAATGGA 1380
DB 1321 GCAGAGGTAGAAATCCACGCAAGTAAAGAAATTTCTTGGAAAGAAACACGCAACCAATGGA 1380

QY 1381 ATTAATTAACCAATAACGAGTGAAGTGTGATTTCCAGAGACAGGAATTAATTTTGGAAAT 1440
DB 1381 ATTAATTAACCAATAACGAGTGAAGTGTGATTTCCAGAGACAGGAATTAATTTTGGAAAT 1440

QY 1441 ACTTTGCTCAGTGGAAATTAAGTATATAAATGACCCCATGCTCCAGAAATTTGTGAAC 1500
DB 1441 ACTTTGCTCAGTGGAAATTAAGTATATAAATGACCCCATGCTCCAGAAATTTGTGAAC 1500

QY 1501 ATAGTAGAGCCAAATGGTAGGATTTATAGGGTTTCCAGAGAGATGCTGCTGACCTTCA 1560
DB 1501 ATAGTAGAGCCAAATGGTAGGATTTATAGGGTTTCCAGAGAGATGCTGCTGACCTTCA 1560

QY 1561 TCACTTGTCTACATTTATATATAACAGCCATCCAGATAAATACTTCTCCAGAGGAATAGA 1620
DB 1561 TCACTTGTCTACATTTATATATAACAGCCATCCAGATAAATACTTCTCCAGAGGAATAGA 1620

QY 1621 GCTGAAGACCACTTCCAGCTTCTTACTACAGGAATTCAGTATATTTACTGACATGCA 1680
DB 1621 GCTGAAGACCACTTCCAGCTTCTTACTACAGGAATTCAGTATATTTACTGACATGCA 1680

QY 1681 GCTGAACAGCTGCTTATGTTATCAGGAGACTTGTACCTTTTCACTGAGCAGCATGATGAT 1740
DB 1681 GCTGAACAGCTGCTTATGTTATCAGGAGACTTGTACCTTTTCACTGAGCAGCATGATGAT 1740

QY 1741 GTATCTGCTTTCACATCATGAGGAGGAGCAACCAACAGTTTCTCAGTGTGAGGACACGT 1800
DB 1741 GTATCTGCTTTCACATCATGAGGAGGAGCAACCAACAGTTTCTCAGTGTGAGGACACGT 1800

QY 1801 CAGCAAGTGCAGAGCTCAATTAATAATTAATAAATAATAAATAATTAAGTCTTCAATCTATT 1860
DB 1801 CAGCAAGTGCAGAGCTCAATTAATAATTAATAAATAATAAATAATTAAGTCTTCAATCTATT 1860
QY 1861 TTGTTATATTGGGATCCTCCAGATATCCCAATGGAAATAAATCACTCACTATACGATTTAT 1920
DB 1861 TTGTTATATTGGGATCCTCCAGATATCCCAATGGAAATAAATCACTCACTATACGATTTAT 1920
QY 1921 GCAATGGAATTTGGATACAAACAGAGCAATTCAGATAAATCACTACATAGATAACAGCTTCTC 1980
DB 1921 GCAATGGAATTTGGATACAAACAGAGCAATTCAGATAAATCACTACATAGATAACAGCTTCTC 1980
QY 1981 ATACACAGGTTAAAGAAATACACAAATACAAATAGAGTGGCAGGCTCAACCCAGAT 2040
DB 1981 ATACACAGGTTAAAGAAATACACAAATACAAATAGAGTGGCAGGCTCAACCCAGAT 2040
QY 2041 GGAGAAAGTTCTTGTCTGAGAAATGACATCTTTGTGAGAACTTCAGAAAGATGAACCG 2100
DB 2041 GGAGAAAGTTCTTGTCTGAGAAATGACATCTTTGTGAGAACTTCAGAAAGATGAACCG 2100
QY 2101 GAATCATCACCTCAAGATGTGGAAGTAAATGATGTTACCCAGATGAAATAAGGTTGAAG 2160
DB 2101 GAATCATCACCTCAAGATGTGGAAGTAAATGATGTTACCCAGATGAAATAAGGTTGAAG 2160
QY 2161 TGGTCAACCCGAAAGCCCAATGGGATCAATTAATGCTTATGAAGTGTCTATATAAAAT 2220
DB 2161 TGGTCAACCCGAAAGCCCAATGGGATCAATTAATGCTTATGAAGTGTCTATATAAAAT 2220
QY 2221 ATAGATACTTTATATAGAGACACATCAACACAGACATATATAAAGAACTTAAGA 2280
DB 2221 ATAGATACTTTATATAGAGACACATCAACACAGACATATATAAAGAACTTAAGA 2280
QY 2281 CCTCACACCTCTATACATTTCTGTAAGGTCTTACACAGATTTGGTCAATGCAATCAG 2340
DB 2281 CCTCACACCTCTATACATTTCTGTAAGGTCTTACACAGATTTGGTCAATGCAATCAG 2340
QY 2341 GTATCTTCTTACTCTCTGTAAGGACTTCGGAGACTGTGCTGATAGTGCCACAGAAAT 2400
DB 2341 GTATCTTCTTACTCTCTGTAAGGACTTCGGAGACTGTGCTGATAGTGCCACAGAAAT 2400
QY 2401 ATCACTTACAAATAATTTCTCTGGAGAGATTTGAGCTATCATTTCTTCCCTCCCAAGTGT 2460
DB 2401 ATCACTTACAAATAATTTCTCTGGAGAGATTTGAGCTATCATTTCTTCCCTCCCAAGTGT 2460
QY 2461 CCCAATGGAATCAATAAATAATACAAATTTATCTCAAGAGAAATGGAATGAGAA 2520
DB 2461 CCCAATGGAATCAATAAATAATACAAATTTATCTCAAGAGAAATGGAATGAGAA 2520
QY 2521 AGAACTATAAATCAACCTCTTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAA 2580
DB 2521 AGAACTATAAATCAACCTCTTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAA 2580
QY 2581 TATATCACTGAGGTGCTGTAGTACACTGAAGGTGAAGGATTCGGAGTGTCCCATATA 2640
DB 2581 TATATCACTGAGGTGCTGTAGTACACTGAAGGTGAAGGATTCGGAGTGTCCCATATA 2640
QY 2641 AGTATCTGACGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAGACAG 2700
DB 2641 AGTATCTGACGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAGACAG 2700
QY 2701 TTGTCTGGTGTACGGTGAAGTGTCTCATGGCAACCCCTGGAGCCAAATGGAATATC 2760
DB 2701 TTGTCTGGTGTACGGTGAAGTGTCTCATGGCAACCCCTGGAGCCAAATGGAATATC 2760
QY 2761 CTTTATACAGTTTATGTCTGGAATAGATCAATTAATAAATAATTAATGTCATGAA 2820
DB 2761 CTTTATACAGTTTATGTCTGGAATAGATCAATTAATAAATAATTAATGTCATGAA 2820
QY 2821 ACATCATGAGTTATCAGATTTGGATTAATAATGTTGAATACAGTGTCTATGTAACAGCT 2880
DB 2821 ACATCATGAGTTATCAGATTTGGATTAATAATGTTGAATACAGTGTCTATGTAACAGCT 2880

QY 2881 AGCACAGATTTGGTGAATGGAAAAACAGGAAGCAATATCAATTAAGCTTTCAAAACACACAGAG 2940
DB 2881 AGCACAGATTTGGTGAATGGAAAAACAGGAAGCAATATCAATTAAGCTTTCAAAACACACAGAG 2940
QY 2941 GGAGACCAAGCGATCCTCCCAAGATGTTTATATGCAAACTCAGTTCCTTCATCAATA 3000
DB 2941 GGAGACCAAGCGATCCTCCCAAGATGTTTATATGCAAACTCAGTTCCTTCATCAATA 3000
QY 3001 ATTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTTATACAAATTAATTAATCTCTGTTAT 3060
DB 3001 ATTCTTTCTGGACACCTCTTCAAAACCTTAATGGGATTTATACAAATTAATTAATCTCTGTTAT 3060
QY 3061 TACAGAAATACCTTCAGGTACTTTTATGCAAAATTTTACCTCCATGCAATTAACCAATGAC 3120
DB 3061 TACAGAAATACCTTCAGGTACTTTTATGCAAAATTTTACCTCCATGCAATTAACCAATGAC 3120
QY 3121 TTTGACAAATGACTGTATCCCAATATAGATAAATGACATATTTGAGTCTACTATACA 3180
DB 3121 TTTGACAAATGACTGTATCCCAATATAGATAAATGACAAATTTGAGTCTACTATACA 3180
QY 3181 TTTTGGTTAAACAGAACTACTTCAGTTGGAAATGGGAAATAAAGACAGTGACATCAATGAA 3240
DB 3181 TTTTGGTTAAACAGAACTACTTCAGTTGGAAATGGGAAATAAAGACAGTGACATCAATGAA 3240
QY 3241 GTATACACAGATCAAGACATACCTGAAGGTTTGTGGAAACCTGACTTACGAATCCATT 3300
DB 3241 GTATACACAGATCAAGACATACCTGAAGGTTTGTGGAAACCTGACTTACGAATCCATT 3300
QY 3301 TCGTCAACTGCAATTAATGTAAGCTGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360
DB 3301 TCGTCAACTGCAATTAATGTAAGCTGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360
QY 3361 TACTATGTTTCACTGATCTTACAGACAGCTCTCGCAATGTGAGACACCTCTGTTTACA 3420
DB 3361 TACTATGTTTCACTGATCTTACAGACAGCTCTCGCAATGTGAGACACCTCTGTTTACA 3420
QY 3421 TATGAGAGAAAGCATATTTTGTATATCTGAAAAATACACTGATTAATATTTAAAAAT 3480
DB 3421 TATGAGAGAAAGCATATTTTGTATATCTGAAAAATACACTGATTAATATTTAAAAAT 3480
QY 3481 ACTCCATCAACAGAAAGGATTTCTGATACCTATACCTGATCTGCCAGCTATACATCAAGCT 3540
DB 3481 ACTCCATCAACAGAAAGGATTTCTGATACCTATACCTGATCTGCCAGCTATACATCAAGCT 3540
QY 3541 GAAGAAGATGTCACAGAACTTCAACCAATTAACCAATTAACCAATTAACCAATTAACCAAT 3600
DB 3541 GAAGAAGATGTCACAGAACTTCAACCAATTAACCAATTAACCAATTAACCAATTAACCAAT 3600
QY 3601 TCAGTTCTCTTATCAATGGGATCCCAAGTAAAGCCAAATGGTGCATATAAATAGTTATGAT 3660
DB 3601 TCAGTTCTCTTATCAATGGGATCCCAAGTAAAGCCAAATGGTGCATATAAATAGTTATGAT 3660
QY 3661 TTAACTTTACAGGACCAATGAAATTAATTTCTTCTTCAATTTCTGATTAATTAATTAATA 3720
DB 3661 TTAACTTTACAGGACCAATGAAATTAATTTCTTCTTCAATTTCTGATTAATTAATTAATA 3720
QY 3721 TTGGAAGAGCTTTCACCTTATATAGCTTTTGTGTCGCCCAAGAACTAGAAA 3780
DB 3721 TTGGAAGAGCTTTCACCTTATATAGCTTTTGTGTCGCCCAAGAACTAGAAA 3780
QY 3781 GGACTTGGTCTTCCAGTATCTTTTACAGATGAGTCAAGTGCCTGTTAGCACCT 3840
DB 3781 GGACTTGGTCTTCCAGTATCTTTTACAGATGAGTCAAGTGCCTGTTAGCACCT 3840
QY 3841 CCACAAATTTGACTTAATCACTGTACTTTCAGCTTTTGTATGGTGAATGAGGCCA 3900
DB 3841 CCACAAATTTGACTTAATCACTGTACTTTCAGCTTTTGTATGGTGAATGAGGCCA 3900
QY 3901 AGTCTCTTCCAGGTGTTTAAAGTATATAGTTTAAATTCATGAACATGAACCT 3960
DB 3901 AGTCTCTTCCAGGTGTTTAAAGTATATAGTTTAAATTCATGAACATGAACCT 3960
QY 3961 GACACTATATATATAAGAAATATATCAGGATTTAAAACTGAAGCCAAACTTGTGTGACTG 4020


```
Qy 6181 AACATAAAACATATAATAATAAACAAGTAAAGCTGATAGCTGACGCTAGTGTTC 6240
Dy 6181 AACATAAAACATATAATAATAAACAAGTAAAGCTGATAGCTGACGCTAGTGTTC 6240
Qy 6241 GGTTCGATATATTAATGCGAGTATATTTCTGTTATTTATGTCCTCAATGAATTTAT 6300
Dy 6241 GGTTCGATATATTAATGCGAGTATATTTCTGTTATTTATGTCCTCAATGAATTTAT 6300
Qy 6301 GCTACTCAAGGTCCACTACCAAGCAAGTGGAGATTTTGGAGATGGTGTGGGAAACC 6360
Dy 6301 GCTACTCAAGGTCCACTACCAAGCAAGTGGAGATTTTGGAGATGGTGTGGGAAACC 6360
Qy 6361 AGGCAAAACATTAATGATGCTAAACAAGTGTGAAAAAGGACGATCAGATGCCAT 6420
Dy 6361 AGGCAAAACATTAATGATGCTAAACAAGTGTGAAAAAGGACGATCAGATGCCAT 6420
Qy 6421 CAGTATGGCCAGAGACACAGCCAGTACTGCTCTTGGAGATATAGTATCAAG 6480
Dy 6421 CAGTATGGCCAGAGACACAGCCAGTACTGCTCTTGGAGATATAGTATCAAG 6480
Qy 6481 CTAATGAGGATGTTCAATAGATTGACTATCAGGATCTGAAATTTGAAAGCGCATGG 6540
Dy 6481 CTAATGAGGATGTTCAATAGATTGACTATCAGGATCTGAAATTTGAAAGCGCATGG 6540
Qy 6541 GATTGCAATGCTGTTGCAAGTGTAACTTTACTGCTGCGCAGAGCATGGGGTTCCTGAG 6600
Dy 6541 GATTGCAATGCTGTTGCAAGTGTAACTTTACTGCTGCGCAGAGCATGGGGTTCCTGAG 6600
Qy 6601 AACAGCCCTCTAATCTACTTTGCAAGTGTGAGTGTGAGCAAGCGGACATGACACC 6660
Dy 6601 AACAGCCCTCTAATCTACTTTGCAAGTGTGAGTGTGAGCAAGCGGACATGACACC 6660
Qy 6661 ACACCTATGATGTTCACTCAGTGTGAGTGTGAGCAAGCGGACATGACACC 6720
Dy 6661 ACACCTATGATGTTCACTCAGTGTGAGTGTGAGCAAGCGGACATGACACC 6720
Qy 6721 GACCATTTAAACACAATATAAATGACCATGATTTTGTGATATATATGACATAGCT 6780
Dy 6721 GACCATTTAAACACAATATAAATGACCATGATTTTGTGATATATATGACATAGCT 6780
Qy 6781 GAATGAGAGTGAAGATGTCAGTGTGAGTGTGAGCAAGTGTGACATATATTTTACAC 6840
Dy 6781 GAATGAGAGTGAAGATGTCAGTGTGAGTGTGAGCAAGTGTGACATATATTTTACAC 6840
Qy 6841 CAGTCAATCTGATCTCTTATCAATAAGGGAAGTAAATCAGCCCATCTGTTTGTAAAC 6900
Dy 6841 CAGTCAATCTGATCTCTTATCAATAAGGGAAGTAAATCAGCCCATCTGTTTGTAAAC 6900
Qy 6901 TATTGACACTTCAGAGATGGACTCTTTGGAGCCCATGGAAGTGTGAGCTTGA 6960
Dy 6901 TATTGACACTTCAGAGATGGACTCTTTGGAGCCCATGGAAGTGTGAGCTTGA 6960
Qy 6961 TGGAGAAACCACTATGTAATATTAGACCAAGGATCAATGGAAGATTTTAA 7020
Dy 6961 TGGAGAAACCACTATGTAATATTAGACCAAGGATCAATGGAAGATTTTAA 7020
Qy 7021 ATCCAGGGCCAAAGTATACCCCTCATTTCTCGAATGGAATGCAACCTTAAGAA 7080
Dy 7021 ATCCAGGGCCAAAGTATACCCCTCATTTCTCGAATGGAATGCAACCTTAAGAA 7080
Qy 7081 ATATCTATGCTTCTCTCACTGTGCTTT 7108
Dy 7081 ATATCTATGCTTCTCTCACTGTGCTTT 7108
```

RESULT 2

AAD56411

ID AAD56411 standard; DNA; 6897 BP.

XX

AC AAD56411;

XX

DT 07-AUG-2003 (first entry)

XX

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Dy

Human receptor tyrosine kinase DNA #1.

Human; receptor tyrosine phosphatase; diabetes; obesity; CNS disorder;
multiple sclerosis; epilepsy; chronic obstructive pulmonary disease;
COPD; Parkinson's disease; stroke; cardiovascular disorder; arrhythmia;
congestive heart failure; myocardial infarction; ischaemic heart disease;
gene therapy; anorectic; cardiast; neuroprotective; anticonvulsant;
cerebroprotective; vasotropic; antiarrhythmic; receptor; enzyme; gene;
chromosome 12; ds.

Homo sapiens.

Key Location/Qualifiers

CDS 1..6897

/tag= a

/product= "Human receptor tyrosine phosphatase protein"

/note= "No stop codon"

/partial

W02003033688-A1.

24-APR-2003.

14-OCT-2002; 2002WO-BP011473.

16-OCT-2001; 2001US-0329329P.

(FARB) BAYER AG.

Liou J;

WPI; 2003-403215/38.

P-PSDB; AAB37320.

Novel polynucleotides encoding human receptor tyrosine phosphatase
polypeptides, useful for treating diabetes, CNS disorders, obesity,
chronic obstructive pulmonary disease and cardiovascular disorders.

Claim 1; Fig 1; 163pp; English.

The present invention relates to receptor tyrosine phosphatase proteins
and polynucleotides encoding them. Sequences of the invention are useful
in the preparation of medicaments for modulating the activity of receptor
tyrosine phosphatase in disease such as diabetes, obesity, CNS disorders
(multiple sclerosis, epilepsy, Parkinson's disease and stroke), chronic
obstructive pulmonary disease (COPD) and cardiovascular disorders (e.g.
congestive heart failure, myocardial infarction, ischaemic heart disease
and arrhythmia). They are also used in gene therapy. The present sequence
is human receptor tyrosine kinase DNA. Receptor tyrosine kinase gene is
located on chromosome 12

Sequence 6897 BP; 2322 A; 1365 C; 1296 G; 1914 T; 0 U; 0 Other;

Query Match

Best Local Similarity 95.7%; Score 6800.8; DB 9; Length 6897;

Matches 6858; Conservative 0; Mismatches 12; Indels 30; Gaps 2;

Qy 106 ATGGATTTTCTTATCATTTTCTTTTCTTTTATTTGGGACTTCAGAGACAGGTTGAT 165

Dy 1 ATGGATTTTCTTATCATTTTCTTTTCTTTTATTTGGGACTTCAGAGACAGGTTGAT 60

Qy 166 GTTTCCTCAATGTCCTGCTAGTACGATATACCATCTTCAATTTCTACACA 225

Dy 61 GTTTCCTCAATGTCCTGCTAGTACGATATACCATCTTCAATTTCTACACA 120

Qy 226 TACACCTCACCTGTTTACTAGATAAGTACCAAAATGTAACAAACAGGCGCTCCAGTC 285

Dy 121 TACACCTCACCTGTTTACTAGATAAGTACCAAAATGTAACAAACAGGCGCTCCAGTC 180

Qy 286 TTCTAGCGGGGAAAGAGTCTGCTGGGATCTTCTGTTGGGATACACACCT 345

Dy 181 TTCTAGCGGGGAAAGAGTCTGCTGGGATCTTCTGTTGGGATACACACCT 240

QY	346	AATCCAAATCGAAGGATTATATCTTACATTTGTCAAATATAAGGAAGTTTGTCCGTGGATG	405	QY	1426	ATAATTTTGGAAAAATATCTTTGCTCACTGGAAAAATATAGATATATAAATGACCCCATGGCT	1485
DB	241	AATCCAAATGGAAGGATTATATCTTACATTTGTCAAATATAAGGAAGTTTGTCCGTGGATG	300	DB	1321	ATAATTTTGGAAAAATATCTTTGCTCACTGGAAAAATATAGATATATAAATGACCCCATGGCT	1380
QY	406	CAAAACAGTATATACACAAGTCAGATCAAAAGCCACAGACAGCTCTGGAAGTTCTTCTTACTAAT	465	QY	1486	CCAGAAAATTTGTGAACATAGTAGAGCAATGGTAGGATTATATAGAGGTTTCAGACAGATG	1545
DB	301	CAAAACAGTATATACACAAGTCAGATCAAAAGCCACAGACAGCTCTGGAAGTTCTTCTTACTAAT	360	DB	1381	CCAGAAAATTTGTGAACATAGTAGAGCAATGGTAGGATTATATAGAGGTTTCAGACAGATG	1440
QY	466	CTTAATCTCGGAACAACATATGAATTAAGTTTGTCTGTGAAAACAAGTGCTGGCATTTGGA	525	QY	1546	TCGTCTGACCTTCACTCACTTCTGCTACATTTATATTAACAGCCATCCAGATAAAAACCTTT	1605
DB	361	CTTAATCTCGGAACAACATATGAATTAAGTTTGTCTGTGAAAACAAGTGCTGGCATTTGGA	420	DB	1441	TCGTCTGACCTTCACTCACTTCTGCTACATTTATATTAACAGCCATCCAGATAAAAACCTTT	1500
QY	526	GTGTTTAGTGATCCATTTCTCTTCCAAATGTCAGAAAAGTGCTCCAGAAAAGTGTTGAAT	585	QY	1606	CCTGCAAGGAATPAGAGCTGAAGACACAGACTTCAACAGTTGTAACTACAGGAATCAGTAT	1665
DB	421	GTGTTTAGTGATCCATTTCTCTTCCAAATGTCAGAAAAGTGCTCCAGAAAAGTGTTGAAT	480	DB	1501	CCTGCAAGGAATPAGAGCTGAAGACACAGACTTCAACAGTTGTAACTACAGGAATCAGTAT	1560
QY	586	CTCACAGTTGAGGCCCTACAAAGCTTCAGCAGTTTAAGCTGATTTGGTATTACCTCGGCAA	645	QY	1666	ATTACTGACATTTGAGCTGGAACAGCTGCTTATGTTTACAGGAGACTTGTACTTTCACT	1725
DB	481	CTCACAGTTGAGGCCCTACAAAGCTTCAGCAGTTTAAGCTGATTTGGTATTACCTCGGCAA	540	DB	1561	ATTACTGACATTTGAGCTGGAACAGCTGCTTATGTTTACAGGAGACTTGTACTTTCACT	1620
QY	646	CGAAATGGCAAAATTAACAGCTTCAAGATTAGTGTCAACATGCCAGAAAGTGGATAGTA	705	QY	1726	GAGCACATGATTAGTGTATCTGCTTTCAACCATCATGGGGAAGGACCCACCAACAGTTCTC	1785
DB	541	CGAAATGGCAAAATTAACAGCTTCAAGATTAGTGTCAACATGCCAGAAAGTGGATAGTA	600	DB	1621	GAGCACATGATTAGTGTATCTGCTTTCAACCATCATGGGGAAGGACCCACCAACAGTTCTC	1680
QY	706	GTGAAAGATGCTCTCAATCAGAGTAGAGGACATTTTGCATCTGGGAAATTTGCCAGAAATGCAAT	765	QY	1786	AGTGTTAGGACACGTCAGCAAGTCGCAAGCTCCATTTAAATTTATAAACTATAAAAATATT	1845
DB	601	GTGAAAGATGCTCTCAATCAGAGTAGAGGACATTTTGCATCTGGGAAATTTGCCAGAAATGCAAT	660	DB	1681	AGTGTTAGGACACGTCAGCAAGTCGCAAGCTCCATTTAAATTTATAAACTATAAAAATATT	1740
QY	766	GAGAAATAGTGAATCTTTTATGAGGTACAGCCAGCCCTTCTCCAAACCTTTGGTAGAGTT	825	QY	1846	AGTCTTCACTATTTTGTATTATTTGGGATCTCTCCAGAAATATCCCAATGGAAAAATTAACCT	1905
DB	661	GAGAAATAGTGAATCTTTTATGAGGTACAGCCAGCCCTTCTCCAAACCTTTGGTAGAGTT	720	DB	1741	AGTCTTCACTATTTTGTATTATTTGGGATCTCTCCAGAAATATCCCAATGGAAAAATTAACCT	1800
QY	826	ACACCTCCATCGGTTACCAACATTCATCAGCAAGCTTTGACACAGAAATGAGATCAGCTCT	885	QY	1906	CACATATACGATTTTATGCAATGGAAATTTGGATACAAACAGACGATTTCCAGATAACTACCAT	1965
DB	721	ACACCTCCATCGGTTACCAACATTCATCAGCAAGCTTTGACACAGAAATGAGATCAGCTCT	780	DB	1801	CACATATACGATTTTATGCAATGGAAATTTGGATACAAACAGACGATTTCCAGATAACTACCAT	1860
QY	886	GTGTGAAAGAGCCTATCAGTTTGTGTAGTGACACCTTGAGACCTTTATACACATATCTT	945	QY	1966	GATAACAGCTTTCTCATACAGGGTTAAAGAAATACAAAAATACAAAATGAGAGTGGCA	2025
DB	781	GTGTGAAAGAGCCTATCAGTTTGTGTAGTGACACCTTGAGACCTTTATACACATATCTT	840	DB	1861	GATAACAGCTTTCTCATACAGGGTTAAAGAAATACAAAATACAAAATGAGAGTGGCA	1920
QY	946	TTTGAAGTTTTCAGCTGCTACAACTGAAGCAGGTTTATATTGATAGTAGTACGATTTGTCAGAAC	1005	QY	2026	GCCTCAACCCACGATGGGAAAAAGTTCTTTGTCTGTAAGAAAAATGACATCTTTGTGAGAACT	2085
DB	841	TTTGAAGTTTTCAGCTGCTACAACTGAAGCAGGTTTATATTGATAGTAGTACGATTTGTCAGAAC	900	DB	1921	GCCTCAACCCACGATGGGAAAAAGTTCTTTGTCTGTAAGAAAAATGACATCTTTGTGAGAACT	1980
QY	1006	CCAGAAATCAGTGCTGAGAGACCAACCAAAAACCTGGGTAAACAGGCAACATCAGAGAAAG	1065	QY	2086	TCAGAGATGAACCCGGAATCATCACTCAAGATGTCGAAGTAATTTGATGTTTACCCGAGAT	2145
DB	901	CCAGAAATCAGTGCTGAGAGACCAACCAAAAACCTGGGTAAACAGGCAACATCAGAGAAAG	960	DB	1981	TCAGAGATGAACCCGGAATCATCACTCAAGATGTCGAAGTAATTTGATGTTTACCCGAGAT	2040
QY	1066	TCCTTTTCAATTTTATGGGACCCCAACTATATAGTAAACAGGGAATTTTATGTTATAGATT	1125	QY	2146	GAAATAAGGTTGAAGTGGTCAACCCCGAAAAAGCCCAATGGGATCATTTATTCCTTATGAA	2205
DB	961	TCCTTTTCAATTTTATGGGACCCCAACTATATAGTAAACAGGGAATTTTATGTTATAGATT	1020	DB	2041	GAAATAAGGTTGAAGTGGTCAACCCCGAAAAAGCCCAATGGGATCATTTATTCCTTATGAA	2100
QY	1126	GAAATTATGGAACCATCAGTCTGCATTTTGGATPACAGCAACAAAGACCTCAAGTTTGCA	1185	QY	2206	GTGCTATATAAAAAATATAGATATCTTTATATATGAAGAACACATCAACACACACATAATA	2265
DB	1021	GAAATTATGGAACCATCAGTCTGCATTTTGGATPACAGCAACAAAGACCTCAAGTTTGCA	1080	DB	2101	GTGCTATATAAAAAATATAGATATCTTTATATATGAAGAACACATCAACACACACATAATA	2160
QY	1186	TTCACTAAACCTAAACCAATTTAATGTATGTATCTATATTTGGGCTGAAACCAAGTGGCA	1245	QY	2266	TTAAGGAATTAAGGACCTCACACCTCTATATACTTTCTGTAAGGTTTACACCAAGTTT	2325
DB	1081	TTCACTAAACCTAAACCAATTTAATGTATGTATCTATATTTGGGCTGAAACCAAGTGGCA	1140	DB	2161	TTAAGGAATTAAGGACCTCACACCTCTATATACTTTCTGTAAGGTTTACACCAAGTTT	2220
QY	1246	GGGACTGGGCCCAGTCAAAATTTTCAAGTATTCAGTATTCACCTCCACAGATGTTCCAGGGCAGTG	1305	QY	2326	GGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACCTTCGAGAGACTGTGCTGAT	2385
DB	1141	GGGACTGGGCCCAGTCAAAATTTTCAAGTATTCAGTATTCAGTATTCACCTCCACAGATGTTCCAGGGCAGTG	1200	DB	2221	GGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAGGACCTTCGAGAGACTGTGCTGAT	2280
QY	1306	TTTGATTTTCAACTTTCAGAGGTAGAAATCCAGCAAGTAAAGATTACTTGGAGAAACCA	1365	QY	2386	AGTGCAACGAGAAATATCACTTACAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTTC	2445
DB	1201	TTTGATTTTCAACTTTCAGAGGTAGAAATCCAGCAAGTAAAGATTACTTGGAGAAACCA	1260	DB	2281	AGTGCAACGAGAAATATCACTTACAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTTC	2340
QY	1366	CGAACCAAAATGGAATTTTAAACCAATACCGAGTGAAAGTGTAGTTCCAGAGACAGGA	1425	QY	2446	CTTCCCGCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTTATCTCAAGAGAACT	2505
DB	1261	CGAACCAAAATGGAATTTTAAACCAATACCGAGTGAAAGTGTAGTTCCAGAGACAGGA	1320	DB	2341	CTTCCCGCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTTATCTCAAGAGAACT	2400
				QY	2506	AATGGAAATGAGGAAAGAACTATAAATACAACTCTTTTAAACCCCAAAACATTAAGTACTG	2565

Db 2401 AATGGAAATGAGGAAAGCAATATAAATACAACTCTTTAAACCCAAACATATAAGTACTG 2460
Qy
Db 2566 AAGAAATATACCCAAATATATCATTTGAGGTCTCTGCTAGTACACTGAAAGGTGAAGCAGTT 2625
Qy
Db 2461 AAGAAATATACCCAAATATATCATTTGAGGTCTCTGCTAGTACACTCAAGGTGAAGGTT 2520
Qy
Db 2626 CGAGTCTCCCAATAGTATATACGAGGAGAAAGATGCTCTGATTTCTCCCTCAAGAC 2685
Qy
Db 2521 CGAGTCTCCCAATAGTATATACGAGGAGAAAGATGCTCTGATTTCTCCCTCAAGAC 2580
Qy
Db 2686 TTCTCTGTAAGAACAGTTGCTGCTGTGTCACGGTGAAGTTGTCATGGCAACACCCCTGGAG 2745
Qy
Db 2581 TTCTCTGTAAGAACAGTTGCTGCTGTGTCACGGTGAAGTTGTCATGGCAACACCCCTGGAG 2640
Qy
Db 2746 CCAATGGAATATCTCTTTATTAACAGATTTATGCTGGAATAGATCATATATAAAACT 2805
Qy
Db 2641 CCAATGGAATATCTCTTTATTAACAGATTTATGCTGGAATAGATCATATATAAAACT 2700
Qy
Db 2806 ATTAATGTCTACTGAACATCATTTGGAGTTATACAGATTTGGATTAATGTTGAATACAGT 2865
Qy
Db 2701 ATTAATGTCTACTGAACATCATTTGGAGTTATACAGATTTGGATTAATGTTGAATACAGT 2760
Qy
Db 2866 GCTTATGTACAGCTAGCACCAGATTTGGTGTGATGGGAAACACAGAAACATATCATTAGC 2925
Qy
Db 2761 GCTTATGTAAACAGCTAGCACCAGATTTGGTGTGATGGGAAACACAGAAACATATCATTAGC 2820
Qy
Db 2926 TTTCAAACACAGAGGAGCACCAGGATCTCCCAAGATGTTTATATGCAAACTC 2985
Qy
Db 2821 TTTCAAACACAGAGGAGCACCAGGATCTCCCAAGATGTTTATATGCAAACTC 2880
Qy
Db 2986 AGTTCTTCTCAATAATTTCTTTCTGGACACTCTCTTCAAACTAATGGATTAACAA 3045
Qy
Db 2881 AGTTCTTCTCAATAATTTCTTTCTGGACACTCTCTTCAAACTAATGGATTAACAA 2940
Qy
Db 3046 TATTACTCTGTTTATACAGAAATATCTTCAAGGTACTTTTATGTCAGATTTTACATCCAT 3105
Qy
Db 2941 TATTACTCTGTTTATACAGAAATATCTTCAAGGTACTTTTATGTCAGATTTTACATCCAT 3000
Qy
Db 3106 GAACCTAACCAATGACATTTGCAATATGACTGTATCCAAATTAAGTAACTGCAATA 3165
Qy
Db 3001 GAAGTAACCAATGACATTTGCAATATGACTGTATCCAAATTAAGTAACTGCAATA 3060
Qy
Db 3166 TTCAAGTACTATACATTTTGGTTTAAACAGCAAGTACTTCAAGTTGGAATGGGAATAAAGC 3225
Qy
Db 3061 TTCAAGTACTATACATTTTGGTTTAAACAGCAAGTACTTCAAGTTGGAATGGGAATAAAGC 3120
Qy
Db 3226 AGTGACATCATTTGAAATATACAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 3285
Qy
Db 3121 AGTGACATCATTTGAAATATACAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 3180
Qy
Db 3286 ACTTACGAAATCCATTTCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 3345
Qy
Db 3181 ACTTACGAAATCCATTTCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 3240
Qy
Db 3346 AAGGCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACTCTCTGCAATGTGAGA 3405
Qy
Db 3241 AAGGCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACTCTCTGCAATGTGAGA 3300
Qy
Db 3406 CCACCTCTGTTATATATGAGAGAGCATATATTTTGTGATTAATCTGGAATAATACATGAT 3465
Qy
Db 3301 CCACCTCTGTTATATATGAGAGAGCATATATTTTGTGATTAATCTGGAATAATACATGAT 3360
Qy
Db 3466 TATATATTAATAATTAATCTCCATCAACAGAAAGGATTTCTGATACCTATATCTGCCCCAG 3525
Qy
Db 3361 TATATATTAATAATTAATCTCCATCAACAGAAAGGATTTCTGATACCTATATCTGCCCCAG 3420
Qy
Db 3526 CTATACATCAAGACTGAAGAGATGTCACAGAACTTCCACCAATATCAACATTTTAA 3585
Qy
Db 3421 CTATACATCAAGACTGAAGAGATGTCACAGAACTTCCACCAATATCAACATTTTAA 3480
Qy
Db 3586 AACCTTTCTCTACCTCAGTTCTTCTTATCATGCGATCCCCCAAGTAAAGCAAATGCTGCA 3645

Db 3481 AACCTTTCTCTACCTCAGTTCTTCTTATCATGGATCCCCAGTAAAGCCAAATGCTGCA 3540
Qy
Db 3646 ATAATAAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAATTTATTTCTTTCACTTCT 3705
Qy
Db 3541 ATAATAAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAATTTATTTCTTTCACTTCT 3600
Qy
Db 3706 GATAATATACATTAATTTGGAAGAGCTTTCCACCTTTTACATTTATATATATATATATGCTGCC 3765
Qy
Db 3601 GATAATATACATTAATTTGGAAGAGCTTTCCACCTTTTACATTTATATATATATATATGCTGCC 3660
Qy
Db 3766 GCAAGAACTAGAAAAGGACTTGTCTCTCCAGTATTTCTTTTCTTTTACACAGATGAGTCA 3825
Qy
Db 3661 GCAAGAACTAGAAAAGGACTTGTCTCTCCAGTATTTCTTTTCTTTTACACAGATGAGTCA 3720
Qy
Db 3826 GTGCCGTTAGCACCTCCACAAAATTTGACTTTTAACTCAACTGTACTTTCAGACTTTGATGG 3885
Qy
Db 3721 GTGCCGTTAGCACCTCCACAAAATTTGACTTTTAACTCAACTGTACTTTCAGACTTTGATGG 3780
Qy
Db 3886 CTGAAATGGAGCCCAAGTCTCTTCCAGTGTGTTTAAAGTATATATATATATATGTTTAAATTT 3945
Qy
Db 3781 CTGAAATGGAGCCCAAGTCTCTTCCAGTGTGTTTAAAGTATATATATATATATGTTTAAATTT 3840
Qy
Db 3946 CATGAACATGAACCTGACACTATATATTAAGAATATATCAAGATTTTAAACCTGAAGCC 4005
Qy
Db 3841 CATGAACATGAACCTGACACTATATATTAAGAATATATCAAGATTTTAAACCTGAAGCC 3900
Qy
Db 4006 AAACCTTTGGTGGAGTGGAAACCACTACTCTATCCGCTGTATCTCGGTTTCAACAA 4065
Qy
Db 3901 AAACCTTTGGTGGAGTGGAAACCACTACTCTATCCGCTGTATCTCGGTTTCAACAA 3960
Qy
Db 4066 GTTGGAAATGGCAATCAATTTAGTAAATAGTAAATTTTCAACCAAGATTCAGTTTCCA 4125
Qy
Db 3961 GTTGGAAATGGCAATCAATTTAGTAAATAGTAAATTTTCAACCAAGATTCAGTTTCCA 4020
Qy
Db 4126 GATGTCTGTCAGAAATGTCAGTGCATGCGCACTAGCTGCGCACTGCTTTAGTGAATGG 4185
Qy
Db 4021 GATGTCTGTCAGAAATGTCAGTGCATGCGCACTAGCTGCGCACTGCTTTAGTGAATGG 4080
Qy
Db 4186 GATCCACCCAAAAGGCAATTAAGCGAGTATATGTTAAACAGTTGGAAGGAAT 4245
Qy
Db 4081 GATCCACCCAAAAGGCAATTAAGCGAGTATATGTTAAACAGTTGGAAGGAAT 4140
Qy
Db 4246 TCTACAAAAGTTTCTCCCAAGATCACTGTCATTTTCAATAAGCTTCTTGGCAATPACC 4305
Qy
Db 4141 TCTACAAAAGTTTCTCCCAAGATCACTGTCATTTTCAATAAGCTTCTTGGCAATPACC 4200
Qy
Db 4306 TCATATGCTTTTAAAGTAAAGGCTTCAACCTGCTGCTGGAAGTGCATGGAAGCACTGC 4365
Qy
Db 4201 TCATATGCTTTTAAAGTAAAGGCTTCAACCTGCTGCTGGAAGTGCATGGAAGCACTGC 4260
Qy
Db 4366 CATGTGAGCACTACTCTGAAAACAGTTCCAGTGTTCCTCCCAAAATATGCTTTTCTGAT 4425
Qy
Db 4261 CATGTGAGCACTACTCTGAAAACAGTTCCAGTGTTCCTCCCAAAATATGCTTTTCTGAT 4320
Qy
Db 4426 GTTCACTCAACTAGTGCACATTTGACATGGAATGAACCTGACATATCTTCTGCTACTTTT 4485
Qy
Db 4321 GTTCACTCAACTAGTGCACATTTGACATGGAATGAACCTGACATATCTTCTGCTACTTT 4380
Qy
Db 4486 CAAATTTACAAATTTACCACTCAACTGCTGCTCAAAATGCAAAAGATGGGAATCCGA 4545
Qy
Db 4381 CAAATTTACAAATTTACCACTCAACTGCTGCTCAAAATGCAAAAGATGGGAATCCGA 4440
Qy
Db 4546 GAATGTTGAATATCAAAATTTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACA 4605
Qy
Db 4441 GAATGTTGAATATCAAAATTTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACA 4500
Qy
Db 4606 GTATATGGAATTAAGAAATTTAGATGATTAAGATTCAGATGGCTGCCAGCAATGCT 4665
Qy
Db 4501 GTATATGGAATTAAGAAATTTAGATGATTAAGATTCAGATGGCTGCCAGCAATGCT 4560
Qy
Db 4666 GGCATGGAATGCTTCAAACTGGATTTCTACAAAACTTCTGCTGCTGCCCTCCAGATGGT 4725
Qy
Db 4561 GGCATGGAATGCTTCAAACTGGATTTCTACAAAACTTCTGCTGCTGCCCTCCAGATGGT 4620

QY	4726	CCTCCTGAAATGTTT	CATGTAGTAGCAACAT	CACCTTTT	AGCATCAGCATAGCTGGAGT	4785	QY	5806	GAAAGAACGTAAGAT	CATTCTTT	CGTCACTTTT	GTGTATCTTT	CAATAAATTT	CTCTT	5865
DB	4621	CCTCCTGAAATGTTT	CATGTAGTAGCAACAT	CACCTTTT	AGCATCAGCATAGCTGGAGT	4680	DB	5701	GAAAGAACGTAAGAT	CATTCTTT	CGTCACTTTT	GTGTATCTTT	CAATAAATTT	CTCTT	5760
QY	4786	GAACTCTGCTGAT	CATTACTCGGACCAACAT	GTATCTGATTT	GTATCTGATTT	4845	QY	5866	GGAACAGCTATTTT	GTGCAATTT	CGAAGAAATTT	CGAAGAGCAAGAAAGG	TGGCACAATAC		5925
DB	4681	GAACTCTGCTGAT	CATTACTCGGACCAACAT	GTATCTGATTT	GTATCTGATTT	4740	DB	5761	GGAAACAGCTATTTT	GTGCAATTT	CGAAGAAATTT	CGAAGAGCAAGAAAGG	TGGCACAATAC		5820
QY	4846	GATGAATTTAATAT	TATCTCTT	CATCAAGTCAAT	TGAAGAAATTA	4905	QY	5926	TCCTCTCAGGATG	CAGAAATTT	TATTCACACT	TAATTTGAAGCTG	GCATCAGCTCAT	CACAGTG	5985
DB	4741	GATGAATTTAATAT	TATCTCTT	CATCAAGTCAAT	TGAAGAAATTA	4800	DB	5821	TCCTCTCAGGATG	CAGAAATTT	TATTCACACT	TAATTTGAAGCTG	GCATCAGCTCAT	CACAGTG	5880
QY	4906	GATTTAGAAATAT	TCTCAAGAGTAT	TCTGTAGTGA	TCACTGCAATTT	4965	QY	5986	GCAGACCTCGAACT	GAGGACGAGAGAT	TTAAACGCG	-----			6020
DB	4801	GATTTAGAAATAT	TCTCAAGAGTAT	TCTGTAGTGA	TCACTGCAATTT	4860	DB	5881	GCAGACCTCGAACT	GAGGACGAGAGAT	TTAAACGCGT	TACTTAGTTATAGAAATTCATC			5940
QY	4966	GCTGCATATGTAG	AGGGAAGTCAAGT	GCTGAAATG	ATTTGTTACTACTTT	5025	QY	6021	--GCCAATAAGCA	GAATTCCTT	CTCTGCAACAT	GTGTGAAGAGCTTT	TGCAAAACAAAC		6078
DB	4861	GCTGCATATGTAG	AGGGAAGTCAAGT	GCTGAAATG	ATTTGTTACTACTTT	4920	DB	5941	AAGCCAATAAGCA	GAATTCCTT	CTCTGCAACAT	GTGTGAAGAGCTTT	TGCAAAACAAAC		6000
QY	5026	CCAAAGGACCCACT	TAACCAATGAT	CAATTT	CCAGAGATAC	5085	QY	6079	CTAAGATTTCA	GAAGAAATTTT	TCGGAATTT	TACCAAAATTT	CTTCAGGATCTTT	CTTCAACT	6138
DB	4921	CCAAAGGACCCACT	TAACCAATGAT	CAATTT	CCAGAGATAC	4980	DB	6001	CTAAGATTTCA	GAAGAAATTTT	TCGGAATTT	TACCAAAATTT	CTTCAGGATCTTT	CTTCAACT	6060
QY	5086	CAATTAACGTTCT	CTCTCTCTCAACCT	TAATGGAATAT	CCAGTATAT	5145	QY	6139	GATGCTGATCTG	CCCTTTGGAAT	TAGAGCAAAAA	CCGTTTCC	CAAAACAT	ATAAATAT	6198
DB	4981	CAATTAACGTTCT	CTCTCTCTCAACCT	TAATGGAATAT	CCAGTATAT	5040	DB	6061	GATGCTGATCTG	CCCTTTGGAAT	TAGAGCAAAAA	CCGTTTCC	CAAAACAT	ATAAATAT	6117
QY	5146	GTTTACCGAGAGAT	GATCCTACTGCTG	TCAGATTT	CACAACCT	5205	QY	6199	AATAATAACAG	AGTAAAGCTG	ATAGTGA	CGCTAGTGT	TCAGGTT	CGGATTTATTAAT	6258
DB	5041	GTTTACCGAGAGAT	GATCCTACTGCTG	TCAGATTT	CACAACCT	5100	DB	6118	AATAATAACAG	AGTAAAGCTG	ATAGTGA	CGCTAGTGT	TCAGGTT	CGGATTTATTAAT	6177
QY	5206	ACCAACACATTTG	CTATTCGAATCT	AGAAAGCT	TAATAAGGT	5265	QY	6259	GCAGCTATTTT	CTGTTTATTT	TATGTC	CAAAATGAATTT	TATTTGCT	TACTCAAGCT	6318
DB	5101	ACCAACACATTTG	CTATTCGAATCT	AGAAAGCT	TAATAAGGT	5160	DB	6178	GCAGCTATTTT	CTGTTTATTT	TATGTC	CAAAATGAATTT	TATTTGCT	TACTCAAGCT	6237
QY	5266	AGTGTTTACG	AGTCAATAGT	CTGCTG	CGAGTCC	5325	QY	6319	CCAGGAAACAG	TGTTGGAGAA	TGTTGG	AGAAAC	CCAGGCA	AAAAACAT	6378
DB	5161	AGTGTTTACG	AGTCAATAGT	CTGCTG	CGAGTCC	5220	DB	6238	CCAGGAAACAG	TGTTGGAGAA	TGTTGG	AGAAAC	CCAGGCA	AAAAACAT	6297
QY	5326	GATATCAAAAGCT	CCAGCAGCA	CAAAACCA	ACCCCTAT	5385	QY	6379	ATGCTAAACAG	CTGTTTGG	AAAAAGG	ACGATC	AGATG	CGCATATTTGGC	6438
DB	5221	GATATCAAAAGCT	CCAGCAGCA	CAAAACCA	ACCCCTAT	5280	DB	6298	ATGCTAAACAG	CTGTTTGG	AAAAAGG	ACGATC	AGATG	CGCATATTTGGC	6357
QY	5386	AAACTGCTTGTG	ACTTCAACAAAT	TACAAT	TGCAATG	5445	QY	6439	AAACAGCCAG	TGTTTCTT	CGGAGAT	TAGTGT	TACAAAGCT	TAATGGAGAT	6498
DB	5281	AAACTGCTTGTG	ACTTCAACAAAT	TACAAT	TGCAATG	5340	DB	6358	AAACAGCCAG	TGTTTCTT	CGGAGAT	TAGTGT	TACAAAGCT	TAATGGAGAT	6417
QY	5446	GATCATGGA	CCCAATTAATAAT	GTACAAAGT	CTTGCGACAGAA	5505	QY	6499	ATAGATTTG	GACTATCAG	GGATCTG	AAAAAT	TGAAAGG	CAATGGG	6558
DB	5341	GATCATGGA	CCCAATTAATAAT	GTACAAAGT	CTTGCGACAGAA	5400	DB	6418	ATAGATTTG	GACTATCAG	GGATCTG	AAAAAT	TGAAAGG	CAATGGG	6477
QY	5506	GGAAATGTAA	CAAGTGGTAT	GATGATAT	TTTAAAT	5565	QY	6559	CAGTGTAAC	CTTTTAC	TGCGCC	CAGAGCA	TGCGGGT	CTCTG	6618
DB	5401	GGAAATGTAA	CAAGTGGTAT	GATGATAT	TTTAAAT	5460	DB	6478	CAGTGTAAC	CTTTTAC	TGCGCC	CAGAGCA	TGCGGGT	CTCTG	6537
QY	5566	GAAAGCTTTCT	TAAACCTCAT	GTAAGAA	GACAAAGT	5625	QY	6619	CACCTTGTG	GAAGTGGT	TCGAG	CAAGC	AGGCA	CATGAC	6678
DB	5461	GAAAGCTTTCT	TAAACCTCAT	GTAAGAA	GACAAAGT	5520	DB	6538	CACCTTGTG	GAAGTGGT	TCGAG	CAAGC	AGGCA	CATGAC	6597
QY	5626	ATCTACAT	CATAGGTC	GTATAT	GTGATG	5685	QY	6679	TGCAGTCT	GAGTGG	AGAACT	CGGAGT	TTTTT	TATGCT	6738
DB	5521	ATCTACAT	CATAGGTC	GTATAT	GTGATG	5580	DB	6598	TGCAGTCT	GAGTGG	AGAACT	CGGAGT	TTTTT	TATGCT	6657
QY	5686	AATGGA	CCACTGAA	ACCAAAAG	CAATCT	5745	QY	6739	ATAAATG	ACCATG	ATTTT	TGTGAT	TATATG	GA	6798
DB	5581	AATGGA	CCACTGAA	ACCAAAAG	CAATCT	5640	DB	6658	ATAAATG	ACCATG	ATTTT	TGTGAT	TATATG	GA	6717
QY	5746	GGCAATTTT	ACTGACTCT	GATTTT	CTGACCT	5805	QY	6799	ATGTCAT	GGTGCAG	AACT	CTGGC	ACAGT	ATAT	6858
DB	5641	GGCAATTTT	ACTGACTCT	GATTTT	CTGACCT	5700	DB	6718	ATGTCAT	GGTGCAG	AACT	CTGGC	ACAGT	ATAT	6777
							QY	6859	TTATCAA	ATAAGG	GAAGTA	ATCAG	CCCAT	CTGTTT	6918

Db 6778 TTATCAATAGGAGAGTATACGCCATCTGTTTGTAACTATTGAGCATTGAGAAG 6837
 Qy 6919 ATGGACTCTTTGGAGCCCATGGAAGTGATGTTGAGCTTGAATGGGAAGAAACCACTATG 6978
 Db 6838 ATGGACTCTTTGGAGCCCATGGAAGTGATGTTGAGCTTGAATGGGAAGAAACCACTATG 6897

RESULT 3
 ABA00717
 ID ABA00717 standard; cDNA; 7225 BP.
 XX ABA00717;
 AC ABA00717;
 DT 04-MAR-2003 (first entry)
 XX Human KPP-2 cDNA, Incyte ID No. 7480588CB1.
 XX
 KW Gene; kinase; phosphatase; KPP; cell proliferation; arteriosclerosis;
 KW atherosclerosis; cirrhosis; hepatitis; reproduction; infertility;
 KW paroxysmal nocturnal haemoglobinuria; polycythaemia vera; psoriasis;
 KW primary thrombocytopenia; cancer; development; renal tubular acidosis;
 KW anaemia; mental retardation; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; epilepsy; Sjogren's syndrome; uveitis; asthma;
 KW menstrual cycle; autoimmune; inflammation; rheumatoid arthritis; AIDS;
 KW autoimmune thyroiditis; contact dermatitis; Crohn's disease; allergy;
 KW diabetes mellitus; glomerulonephritis; Goodpasture's syndrome; gout;
 KW Graves' disease; Hashimoto's thyroiditis; irritable bowel syndrome;
 KW multiple sclerosis; osteoarthritis; osteoporosis; pancreatitis;
 KW Reiter's syndrome; rat; glomerular mesangial cell receptor protein;
 KW tyrosine phosphatase precursor; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..6900
 FT /tag= a
 FT /product= "KPP-2"
 XX
 PN WO200290530-A2.
 XX
 PD 14-NOV-2002.
 XX
 PF 16-JAN-2002; 2002WO-US001369.
 XX
 PR 18-JAN-2001; 2001US-0263083P.
 PR 23-FEB-2001; 2001US-0271117P.
 PR 23-FEB-2001; 2001US-0271205P.
 PR 16-MAR-2001; 2001US-0276859P.
 PR 23-MAR-2001; 2001US-0278504P.
 PR 23-MAR-2001; 2001US-0278522P.
 PR 29-MAR-2001; 2001US-0280266P.
 PR 29-MAR-2001; 2001US-0280510P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lee BA, Wallia NK, Baughn MR, Ison CH, Gururajan R, Arvizu C;
 PI Yao MG, Jackson JL, Tang TY, Yue H, Tran B, Ding L, Lu DAM;
 PI Lal PG, Warren BA;
 XX
 XX WPI; 2003-111972/10.
 DR P-PSDB; AAG79724.
 XX
 XX New human kinases and phosphatases and polynucleotides, useful for
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
 PT cancer or hepatitis.
 XX
 XX Claim 5; Page 138-39; 144pp; English.
 PS
 CC The sequences given in ABA00716-23 encode human kinases and phosphatases
 CC (KPP). The KPP polypeptides, polynucleotides, and agonists and
 CC antagonists to them, are useful for diagnosing, treating or preventing

disorders associated with aberrant expression of KPP, particularly cell
 proliferative disorders (e.g. arteriosclerosis, atherosclerosis,
 cirrhosis, hepatitis, paroxysmal nocturnal haemoglobinuria, polycythaemia
 vera, psoriasis, primary thrombocytopenia or cancer), developmental
 disorders (e.g. renal tubular acidosis, anaemia or mental retardation),
 neurological disorders (e.g. Alzheimer's disease, Parkinson's disease or
 epilepsy), reproductive disorders (e.g. infertility or a disruption in
 the menstrual cycle), or autoimmune/inflammatory disorders (e.g. AIDS,
 allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's
 disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,
 gout, Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome,
 multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's
 syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,
 bacterial, fungal, parasitic, protozoan or helminthic infections. They
 are also useful in the assessment of the effects of exogenous compounds
 on the expression of nucleic acid and amino acid sequences of proteins
 associated with KPP. The protein encoded by this sequence is homologous
 to rat glomerular mesangial cell receptor protein tyrosine phosphatase
 precursor
 CC
 XX
 SQ Sequence 7225 BP; 2434 A; 1417 C; 1352 G; 2022 T; 0 U; 0 Other;
 Query Match 95.7%; Score 6799.8; DB 8; Length 7225;
 Best Local Similarity 98.7%; Pred. No. 0;
 Matches 6962; Conservative 0; Mismatches 17; Indels 72; Gaps 8;
 Qy 106 ATGGATTCTTCTATCATCTTTCTTTTCTTTTATTTGGGACTTTCAGAGACACAGGTTGAT 165
 Db 1 ATGGATTCTTCTATCATCTTTCTTTTCTTTTATTTGGGACTTTCAGAGACACAGGTTGAT 60
 Qy 166 GTTTCGAATGTCGTTCTCTGTTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACA 225
 Db 61 GTTTCGAATGTCGTTCTCTGTTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACA 120
 Qy 226 TACACCTCACCTGTTTACTAGATAGTACACCAAAATGTAACAAACACAGGCTCCAGTC 285
 Db 121 TACACCTCACCTGTTTACTAGATAGG-----GAACAGAGGCTCCAGTC 165
 Qy 286 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTGTTGGAAATACACACACCT 345
 Db 166 TTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTCTGTTGGAAATACACACACCT 225
 Qy 346 AATCCAAATGGAAGGATTTATCTTACATGTCAAATATAAGGAAGTTTGTCCGTGGATG 405
 Db 226 AATCCAAATGGAAGGATTTATCTTACATGTCAAATATAAGGAAGTTTGTCCGTGGATG 285
 Qy 406 CAAACAGATATATACCAAGTCAGNTCAAAGCCAGACAGTCTGGAAGTTCTTCTACTAAT 465
 Db 286 CAAACAGATATATACCAAGTCAGNTCAAAGCCAGACAGTCTGGAAGTTCTTCTACTAAT 345
 Qy 466 CTTAATCTGGAAACACATATGAAATTAAGGTTGCTGCTGAAACAGTCTGGCATTGGA 525
 Db 346 CTTAATCTGGAAACACATATGAAATTAAGGTTGCTGCTGAAACAGTCTGGCATTGGA 405
 Qy 526 GTGTTTAGTGATCCATTTCTTCTCCAACTGCAGAAAGTGCTCCAGGAAAGTGGTGAAT 585
 Db 406 GTGTTTAGTGATCCATTTCTTCTCCAACTGCAGAAAGTGCTCCAGGAAAGTGGTGAAT 465
 Qy 586 CTCACAGTTGAGCCCTACAAAGCTTCAGCAGTTAAGCTGATTTGGTATTACCTCGGCAA 645
 Db 466 CTCACAGTTGAGCCCTACAAAGCTTCAGCAGTTAAGCTGATTTGGTATTACCTCGGCAA 525
 Qy 646 CCAAAATGGCAAAATTACCAAGTTCAAGATTAGTGTCAAATGSCCAGAAAGTGGATAGTA 705
 Db 526 CCAAAATGGCAAAATTACCAAGTTCAAGATTAGTGTCAAATGSCCAGAAAGTGGATAGTA 585
 Qy 706 GTGAAGAGTGTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTCGCAGAAATGCAAT 765
 Db 586 GTGAAGAGTGTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTCGCAGAAATGCAAT 645
 Qy 766 GAGATAGTGAATCTTTTATTTAGGAGTACAGCCAGCCCTTCTCCAAACCTTGTGATGTT 825
 Db 646 GAGATAGTGAATCTTTTATTTAGGAGTACAGCCAGCCCTTCTCCAAACCTTGTGATGTT 705

QY	826	ACACCTCCATCGGTAACCAACATTCATCAAGCAGCGTTGACACAGAAATGAGATCAGCTCT	885	QY	1900	ATAAATCACTACATACGATTTATGCAATGGAATTTGGATACAAACAGAGCAATTCAGATAAAT	1959
DB	706	ACACCTCCATCGGTAACCAACATTCATCAAGCAGCGTTGACACAGAAATGAGATCAGCTCT	765	DB	1783	ATAAATCACTACATACGATTTATGCAATGGAATTTGGATACAAACAGAGCAATTCAGATAAAT	1842
QY	886	GTGTGGAAGAGAGCTATCAGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	945	QY	1960	ACCATAGATTAACAGCTTTCTCTATAACAG-----GGTTAAAGAAATACACAAA	2007
DB	766	GTGTGGAAGAGAGCTATCAGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	825	DB	1843	ACCATAGATTAACAGCTTTCTCTATAACAG-----GGTTAAAGAAATACACAAA	1902
QY	946	TTTGAAGTTTCTAGCTGCTCAACTGAAAGCAGGTTATTTGATAGTACGATTTGTTCAGAAAC	1005	QY	2008	TACAAAATGAGAGTGGCAGCCTCAACCCACAGATGAGAGAAATTTCTTGTCTGAGAAAT	2067
DB	826	TTTGAAGTTTCTAGCTGCTCAACTGAAAGCAGGTTATTTGATAGTACGATTTGTTCAGAAAC	885	DB	1903	TACAAAATGAGAGTGGCAGCCTCAACCCACAGTGGAGAAATTTCTTGTCTGAGAAAT	1962
QY	1006	CCAGATCAGTGGCTGAAGGACCAACCAAACTGCTGTAACAGGCAACATCAACAGAAAG	1065	QY	2068	GACATCTTTGTGAGAACTTCAGAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTA	2127
DB	886	CCAGATCAGTGGCTGAAGGACCAACCAAACTGCTGTAACAGGCAACATCAACAGAAAG	945	DB	1963	GACATCTTTGTGAGAACTTCAGAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTA	2022
QY	1066	TCCTTTTCAATTTTATGGGACCCCAACCACTATAGTAAACAGGGAAATTTTATAGTTATAGATT	1125	QY	2128	ATTGATGTTTACCGCAGATGAATAAAGTTGAAAGTGGTCAACCCGGAAGGCCAATGGG	2187
DB	946	TCCTTTTCAATTTTATGGGACCCCAACCACTATAGTAAACAGGGAAATTTTATAGTTATAGATT	1005	DB	2023	ATTGATGTTTACCGCAGATGAATAAAGTTGAAAGTGGTCAACCCGGAAGGCCAATGGG	2082
QY	1126	GAAATATATGGACCAT-----CAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAG	1179	QY	2188	ATCATTTATGCTTATGAGTGTCTATATAAATAATATAGATCTTTATATATGAGAACACA	2247
DB	1006	GAAATATATGGACCAT-----CAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAG	1065	DB	2083	ATCATTTATGCTTATGAGTGTCTATATAAATAATATAGATCTTTATATATGAGAACACA	2142
QY	1180	TTTGCATTTCACTTAACCTTAACCACTTTAATGATGATGATGATGATGATGATGATGATGATG	1239	QY	2248	TCACAAACAGACATTAATTTAAGGAACTTTAAGACCTTCACACCTCTATATAAATTTCTGTA	2307
DB	1066	TTTGCATTTCACTTAACCTTAACCACTTTAATGATGATGATGATGATGATGATGATGATGATG	1125	DB	2143	TCACAAACAGACATTAATTTAAGGAACTTTAAGACCTTCACACCTCTATATAAATTTCTGTA	2202
QY	1240	AGTGACGAGCTGGGCCCCAAGTCAAAATATTTCACTGATTTCACTCCACAGATGTTCCAGGG	1299	QY	2308	AGGTCTTACACAGATTTTGGTCAATGAGTATCTTTTACTCTCTCTCTCTCTCTCTCTCTCTCT	2367
DB	1126	AGTGACGAGCTGGGCCCCAAGTCAAAATATTTCACTGATTTCACTCCACAGATGTTCCAGGG	1185	DB	2203	AGGTCTTACACAGATTTTGGTCAATGAGTATCTTTTACTCTCTCTCTCTCTCTCTCTCTCTCT	2262
QY	1300	GCAAGTGTGTTGATTTCAACTTGACAGGTAGAAATCCACGCAAGTAAGAAATTAATTCTGGAAG	1359	QY	2368	TCGGAGACTGTGCTGATAGTGCACACAGAAATATACACTTACAAAATATTTCTTCTGGA	2427
DB	1186	GCAAGTGTGTTGATTTCAACTTGACAGGTAGAAATCCACGCAAGTAAGAAATTAATTCTGGAAG	1245	DB	2263	TCGGAGACTGTGCTGATAGTGCACACAGAAATATACACTTACAAAATATTTCTTCTGGA	2322
QY	1360	AAACACGACCAACCAATGGAATTTAACCATAACCGAGTGAAGTGTCTGTCAGAG	1419	QY	2428	GAGATTGAGCTATCACTTCTCTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACA	2487
DB	1246	AAACACGACCAACCAATGGAATTTAACCATAACCGAGTGAAGTGTCTGTCAGAG	1305	DB	2323	GAGATTGAGCTATCACTTCTCTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACA	2382
QY	1420	ACAGGAATTAATTTTGAAATATCTTGTCTCACTGGAATAATGAGTATATAATGACCC	1479	QY	2488	ATTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATATAAATAAATCTTTTAAAC	2547
DB	1306	ACAGGAATTAATTTTGAAATATCTTGTCTCACTGGAATAATGAGTATATAATGACCC	1362	DB	2383	ATTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATATAAATAAATCTTTTAAAC	2442
QY	1480	ATGGCTCCAGAAATGTTGAACATAGTAGAGCCCAATGGTAGGATATATGAGGTTTCAGCA	1539	QY	2548	CAAAACATTAAGTACTGAGAAATATACCAATATATCATTTGAGGTGTCTGCTAGTACA	2607
DB	1363	ATGGCTCCAGAAATGTTGAACATAGTAGAGCCCAATGGTAGGATATATGAGGTTTCAGCA	1422	DB	2443	CAAAACATTAAGTACTGAGAAATATACCAATATATCATTTGAGGTGTCTGCTAGTACA	2502
QY	1540	GAGATGCTGCTGACCTTCACTCACTGCTGCTACATTTATATATACAGCCATCCAGATAA	1599	QY	2608	CTGAAAGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATACTGACGAGGAAGATGCTCT	2667
DB	1423	GAGATGCTGCTGACCTTCACTCACTGCTGCTACATTTATATATACAGCCATCCAGATAA	1482	DB	2503	CTCAAAAGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATACTGACGAGGAAGATGCTCT	2562
QY	1600	AACTTTCTGCAAGGAATAGAGCTGAAGACCAAGCTTCCACAGTGTGTAACTACAAAGGAAT	1659	QY	2668	GATTCTCCCTCCAAAGCTTCTGTGTAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2727
DB	1483	AACTTTCTGCAAGGAATAGAGCTGAAGACCAAGCTTCCACAGTGTGTAACTACAAAGGAAT	1542	DB	2563	GATTCTCCCTCCAAAGCTTCTGTGTAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2622
QY	1660	CAGTATATTAATGACATTTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1719	QY	2728	TGGCAACCACTCCCTGGAGCAATGGAATTAATGGAATTAATGGAATTAATGGAATTAATGGA	2784
DB	1543	CAGTATATTAATGACATTTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1602	DB	2623	TGGCAACCACTCCCTGGAGCAATGGAATTAATGGAATTAATGGAATTAATGGAATTAATGGA	2682
QY	1720	TTTCACTGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1779	QY	2785	AATAGATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2844
DB	1603	TTTCACTGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1662	DB	2683	AATAGATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2742
QY	1780	GTCTCAGTGTGAGGACCTGAGCAAGTCCCAAGCTCCATTAATAATTAATAACTATAAA	1839	QY	2845	GATTATTAATGTTGAAATACAGTCTTTATGTAACAGCTAGCAACAGATTTTGGTGAAGAAA	2904
DB	1663	GTCTCAGTGTGAGGACCTGAGCAAGTCCCAAGCTCCATTAATAATTAATAACTATAAA	1722	DB	2743	GATTATTAATGTTGAAATACAGTCTTTATGTAACAGCTAGCAACAGATTTTGGTGAAGAAA	2802
QY	1840	AATATTAGTCTTCACTATTTTGTGTATATTTGGATCTTCCAGAAATATCCCAATGGA	1899	QY	2905	ACAGGAGCAATATCTTACCTTTTCAACACAGAGGAGCACCAGCATCTCTCCCAA	2964
DB	1723	AATATTAGTCTTCACTATTTTGTGTATATTTGGATCTTCCAGAAATATCCCAATGGA	1782	DB	2803	ACAGGAGCAATATCTTACCTTTTCAACACAGAGGAGCACCAGCATCTCTCCCAA	2859

Db 2860 GATGTTTATATGCAACCTCAGTCTCTCATCAATAATCTTTCTGGACACCTCCTTCA 2819
QY 3025 AAACCTAATGGGATATACAAATATTAATCTGTTTATACAGAAATATCTTCAGGTACTTTT 3084
Db 2920 AAACCTAATGGGATATACAAATATTAATCTGTTTATACAGAAATATCTTCAGGTACTTTT 2979
QY 3085 ATGCAGAAATTTTACACTCCATGAACTAAACCAATGACTTTGACAAATATGACTGATCCACA 3144
Db 2980 ATGCAGAAATTTTACACTCCATGAACTAAACCAATGACTTTGACAAATATGACTGATCCACA 3039
QY 3145 ATTATAGATAAATCTGACAAATATTCAGCTACTATACATTTTGGTTTAAACAGCAAGTACTTCA 3204
Db 3040 ATTATAGATAAATCTGACAAATATTCAGCTACTATACATTTTGGTTTAAACAGCAAGTACTTCA 3099
QY 3205 GTTGGAAATGGGAATAAAGCAGTGACATCAATGAAAGTATACACAGATCAAGACATACCT 3264
Db 3100 GTTGGAAATGGGAATAAAGCAGTGACATCAATGAAAGTATACACAGATCAAGACGTAACCT 3159
QY 3265 GAAGGTTTGTGGAAACCTGACTTACGAATCAATTCGTCAACTGCAATTAATGTAAGC 3324
Db 3160 GAAGGTTTGTGGAAACCTGACTTACGAATCAATTCGTCAACTGCAATTAATGTAAGC 3219
QY 3325 TGGGTCCACCGGCTCAACCAAGGCTAGTCTCTACTATGTTTCACTGATCTTACAG 3384
Db 3220 TGGGTCCACCGGCTCAACCAAGGCTAGTCTCTACTATGTTTCACTGATCTTACAG 3279
QY 3385 CAGACTCTCGCCATGTGAGACCACTCTTGTGTACATATGAGAGAGCATATATTTGAT 3444
Db 3280 CAGACTCTCGCCATGTGAGACCACTCTTGTGTACATATGAGAGAGCATATATTTGAT 3339
QY 3445 AATCTGGAATAATACACTGATATATATTAATAATTAATCTCCATCAACAGAAAGGGATTC 3504
Db 3340 AATCTGGAATAATACACTGATATATATTAATAATTAATCTCCATCAACAGAAAGGGATTC 3399
QY 3505 TCTGATACCTTACTGCCAGCTATACATCAAGCTGAAGAGATGTCACAGAACTTCA 3564
Db 3400 TCTGATACCTTACTGCCAGCTATACATCAAGCTGAAGAGATGTCACAGAACTTCA 3459
QY 3565 CCAATTAATCAACTTTTAAACCTTTCCTCTACCTCAGTCTCTTATCATGGGATCCC 3624
Db 3460 CCAATTAATCAACTTTTAAACCTTTCCTCTACCTCAGTCTCTTATCATGGGATCCC 3519
QY 3625 CCAGTAAAGCAAATGGTGCATTAATAAGTTATGATTTAACTTTTACAAGGACCAAAATGAA 3684
Db 3520 CCAGTAAAGCAAATGGTGCATTAATAAGTTATGATTTAACTTTTACAAGGACCAAAATGAA 3579
QY 3685 AATTATCTTTCATTTACTCTGATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3744
Db 3580 AATTATCTTTCATTTACTCTGATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3639
QY 3745 TTATATAGCTTTTTCGTCGCCAAGAACTAGAAAAAGGACTTGGTCTCTCCAGTATTTCT 3804
Db 3640 TTATATAGCTTTTTCGTCGCCAAGAACTAGAAAAAGGACTTGGTCTCTCCAGTATTTCT 3699
QY 3805 TTCTTTTACACAGATGAGTGCCTTTAGCACCTCCACAAATTTGACTTTTAAATCAAC 3864
Db 3700 TTCTTTTACACAGATGAGTGCCTTTAGCACCTCCACAAATTTGACTTTTAAATCAAC 3759
QY 3865 TGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGTATGTT 3924
Db 3760 TGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGTATGTT 3819
QY 3925 AAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATATATTAATTAAGATATA 3984
Db 3820 AAAGTATATAGTTTAAATTCATGAACATGAACCTGACACTATATATTAATTAAGATATA 3879
QY 3985 TCAGGATTTAAACTGAAGCAAACCTTTGGAGCTGGAACCACTGACACTATATTAATTAAGATATA 4044
Db 3880 TCAGGATTTAAACTGAAGCAAACCTTTGGAGCTGGAACCACTGACACTATATTAATTAAGATATA 3939
QY 4045 CGGTATCTGCTTCAACCAAGTTGGAATGGAATCAATTTAGTAAATTTAGTAAATTTCA 4104

Db 3940 CGGTATCTGCTTCAACCAAGTTGGAATGGAATCAATTTAGTAAATTTAGTAAATTTCA 3999
QY 4105 ACAACCCAAAGAAATCAGTCTCCAGATGTCGTCAGAAATATGCAATGCAATGCAATGCAAT 4164
Db 4000 ACAACCCAAAGAAATCAGTCTCCAGATGTCGTCAGAAATATGCAATGCAATGCAATGCAAT 4059
QY 4165 CAGTCAATTTAGTAAATGGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4224
Db 4060 CAGTCAATTTAGTAAATGGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4119
QY 4225 ATGGTAAACAGTTTCAAAAGGAATTTCTACAAAGTTTCTCCCAAGATCAATGCAATGCAAT 4284
Db 4120 ATGGTAAACAGTTTCAAAAGGAATTTCTACAAAGTTTCTCCCAAGATCAATGCAATGCAAT 4179
QY 4285 ATAAAGCTTTCTGGCAATACCTCATATGTTTAAAGTAAAGCTTTCAACCTCAGCTGGT 4344
Db 4180 ATAAAGCTTTCTGGCAATACCTCATATGTTTAAAGTAAAGCTTTCAACCTCAGCTGGT 4239
QY 4345 GAAGGTGATGAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4404
Db 4240 GAAGGTGATGAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4299
QY 4405 ACAATATTTGCTTTTCTGATGTTTCAAGTCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4464
Db 4300 ACAATATTTGCTTTTCTGATGTTTCAAGTCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4359
QY 4465 GACACTATCTTGGCTACTTTTCAAAATTAACAAATTAACCACTCAACTCTGCTCAAAA 4524
Db 4360 GACACTATCTTGGCTACTTTTCAAAATTAACAAATTAACCACTCAACTCTGCTCAAAA 4419
QY 4525 TGCAAAGAAATGGGAATCCGAAAGATGTTGTTGAATATCAAAATTAACAAATTAACAAATTAACAA 4584
Db 4420 TGCAAAGAAATGGGAATCCGAAAGATGTTGTTGAATATCAAAATTAACAAATTAACAAATTAACAA 4479
QY 4585 GCTCACTTAACCTGAAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4644
Db 4480 GCTCACTTAACCTGAAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4539
QY 4645 GTGGCTGCCAGACCAACTGCTGGCTATGCAATGCTTTCAAACTGGATTTCTACAAAACT 4704
Db 4540 GTGGCTTTCCAGACCAACTGCTGGCTATGCAATGCTTTCAAACTGGATTTCTACAAAACT 4599
QY 4705 GTGGCTGCCCTCCAGATGGTCTCTGAAATGTTTCAATGATGATGATGATGATGATGATGATGAT 4764
Db 4600 GTGGCTGCCCTCCAGATGGTCTCTGAAATGTTTCAATGATGATGATGATGATGATGATGATGAT 4659
QY 4765 AGCATCAGCATGAGTGGAGTGAACCTGCTGCTATCTGGAACCACTGATGATGATGATGATGATGAT 4824
Db 4660 AGCATCAGCATGAGTGGAGTGAACCTGCTGCTATCTGGAACCACTGATGATGATGATGATGATGAT 4719
QY 4825 GATGTCAAAATCGGTAGATAATGATGAATTTAATATATCTTCAATCAAGTCAAAATGAAGAA 4884
Db 4720 GATGTCAAAATCGGTAGATAATGATGAATTTAATATATCTTCAATCAAGTCAAAATGAAGAA 4779
QY 4885 AATAAAACCATAGAAATTAAGATTTAGAAATTAATCAAGGATTTCTGATGATGATGATGATGATGAT 4944
Db 4780 AATAAAACCATAGAAATTAAGATTTAGAAATTAATCAAGGATTTCTGATGATGATGATGATGATGAT 4839
QY 4945 GCATTTACTGGGAACATTAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5004
Db 4840 GCATTTACTGGGAACATTAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4899
QY 5005 GTTACTACTTTTAAAGTCAAGGACCCCAAGGACCCCAAGGACCCCAAGGACCCCAAGGACCCCAAGGAC 5064
Db 4900 GTTACTACTTTTAAAGTCAAGGACCCCAAGGACCCCAAGGACCCCAAGGACCCCAAGGACCCCAAGGAC 4959
QY 5065 CCAGATGAAAGTTTCAAAATTTCAATTTAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5124
Db 4960 CCAGATGAAAGTTTCAAAATTTCAATTTAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5019
QY 5125 ATCCAGATGATATCAAGCTCTGGTTTACCGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 5184
Db 5020 ATCCAGATGATATCAAGCTCTGGTTTACCGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 5079

QY	5185	AACCTCAGTATTATACAGAAACCAACACATTCGTCTATTCGCAATGCTAGAGACTAATA	5244	QY	6238	CCAGGTTCCGATTATTAATTAATGCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTT	6297
DB	5080	AACCTCAGTATTATACAGAAACCAACACATTCGTCTATTCGCAATGCTAGAGACTAATA	5139	DB	6157	CCAGGTTCCGATTATTAATTAATGCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTT	6216
QY	5245	GGTGGACATACATACAAATACAGTGTTTACGCACTCAATAGTGTCTGGTCCAGTCCAAAG	5304	QY	6298	ATTGCTACTCAAGGTCCTACACAGAACAGTTGGAGATTTTGGAGAAATGCTGTGGAA	6357
DB	5140	GGTGGACATACATACAAATACAGTGTTTACGCACTCAATAGTGTCTGGTCCAGTCCAAAG	5199	DB	6217	ATTGCTACTCAAGGTCCTACACAGAACAGTTGGAGATTTTGGAGAAATGCTGTGGAA	6276
QY	5305	GTTCGATGAGAAATACCAATGGAATCAAGTCTCAGCGAGACCAAAACCAAAACCAACC	5364	QY	6358	ACGAGGCAAAACATTAAGTAATGCTAAACACAGTCTTTTGAAGGAGCGATCAGATGC	6417
DB	5200	GTTCGATGAGAAATACCAATGGAATCAAGTCTCAGCGAGACCAAAACCAAAACCAACC	5259	DB	6277	ACCGAGCAAAACATTAAGTAATGCTAAACACAGTCTTTTGAAGGAGCGATCAGATGC	6336
QY	5365	CCTATTATGATGCCACAGGAAATCTGTGTGACTTCAACACAAATTAACATCAGAAAG	5424	QY	6418	CATCAGTATTTGGCAGAGGACCAACAGCCAGTACTGTCTTTGGAGATATAGTGAATCA	6477
DB	5260	CCTATTATGATGCCACAGGAAATCTGTGTGACTTCAACACAAATTAACATCAGAAAG	5319	DB	6337	CATCAGTATTTGGCAGAGGACCAACAGCCAGTACTGTCTTTGGAGATATAGTGAATCA	6396
QY	5425	CCAATATGTTACTACAGTATGATCATGGACCAATAAATAATGTACAAAGTCTTGGACA	5484	QY	6478	AAAGCTAAATGGAGGATGTTCAAAATAGATGGAATCATCAGGATCTGAAATTTGAAAGGCAAT	6537
DB	5320	CCAATATGTTACTACAGTATGATCATGGACCAATAAATAATGTACAAAGTCTTGGACA	5379	DB	6397	AAAGCTAAATGGAGGATGTTCAAAATAGATGGAATCATCAGGATCTGAAATTTGAAAGGCAAT	6456
QY	5485	GAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGCATATTTAATAAA	5544	QY	6538	GGGGAATGCATGACTGTTCCAGCAGTGTAACTTTTACCTGCTGCCAGAGCATGGGTTCT	6597
DB	5380	GAAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGCATATTTAATAAA	5439	DB	6457	GGGGAATGCATGACTGTTCCAGCAGTGTAACTTTTACCTGCTGCCAGAGCATGGGTTCT	6516
QY	5545	GCAAGGCCATATTTTCAAAATGAAGGCTTCTTAACCCCTCCATGTTACAGAGGAAGACA	5604	QY	6598	GAGAACAGCGCCCTCTAATTCACCTTTGTGAAGTGTTCGAGCAAGCAGGCAATGAC	6657
DB	5440	GCAAGGCCATATTTTCAAAATGAAGGCTTCTTAACCCCTCCATGTTACAGAGGAAGACA	5499	DB	6517	GAGAACAGCGCCCTCTAATTCACCTTTGTGAAGTGTTCGAGCAAGCAGGCAATGAC	6576
QY	5605	AGCTTTAGTGGCAATCAAGAAATCTACATCATAGTGTCTGATATGATCATGATGTTCT	5664	QY	6658	ACCACACCTATGATGTTCACTGCACTGTGAGTGTGAAGAACTGGAGTTTTTATGCT	6717
DB	5500	AGCTTTAGTGGCAATCAAGAAATCTACATCATAGTGTCTGATATGATGATGATGTTCT	5559	DB	6577	ACCACACCTATGATGTTCACTGCACTGTGAGTGTGAAGAACTGGAGTTTTTATGCT	6636
QY	5665	GGCAATGAAGACAAAAATTTGCAATGGACCACTGAAACCAAAAGCAATACCTATTATAA	5724	QY	6718	CTGGACCATTTTAAACACATATAATGACCATGATTTTGTGGATATATATGACATGATA	6777
DB	5560	GGCAATGAAGACAAAAATTTGCAATGGACCACTGAAACCAAAAGCAATACCTATTATAA	5619	DB	6637	CTGGACCATTTTAAACACATATAATGACCATGATTTTGTGGATATATATGACATGATA	6696
QY	5725	TTTAGAGCTACAATATATGGACAAATTTACTGCTCTGATATTTCTGACCTGTTAAG	5784	QY	6778	GCTGAACCTGAGAAAGTGAAGAAATGTCATGTCAGAAATCTGGCAGAGTATATCTTTTAA	6837
DB	5620	TTTAGAGCTACAATATATGGACAAATTTACTGCTCTGATATTTCTGACCTGTTAAG	5679	DB	6697	GCTGAACCTGAGAAAGTGAAGAAATGTCATGTCAGAAATCTGGCAGAGTATATCTTTTAA	6756
QY	5785	ACTTTAGGGAAGGACTTTTCAAGAAACCGTAGAGATCAATCTTTCCGTCACCTTTGTGT	5844	QY	6838	CACCACTGCAATCTGATCTCTTATCAAAATAAGGAAGTAATCAGCCCATCTGTTTGT	6897
DB	5680	ACTTTAGGGAAGGACTTTTCAAGAAACCGTAGAGATCAATCTTTCCGTCACCTTTGTGT	5739	DB	6757	CACCACTGCAATCTGATCTCTTATCAAAATAAGGAAGTAATCAGCCCATCTGTTTGT	6816
QY	5845	ATCCTTTCAATAATCTCCTTGGAAACAGCTATTTTGGCAATTTGCAATTCGACAGAAAG	5904	QY	6898	AACTATTCAGCACTTCAGAGATGACTCTTTTGGACCCATGGAAGTGTATGTTAGCTT	6957
DB	5740	ATCCTTTCAATAATCTCCTTGGAAACAGCTATTTTGGCAATTTTGGCAATTTGCAAGAAAG	5799	DB	6817	AACTATTCAGCACTTCAGAGATGACTCTTTTGGACCCATGGAAGTGTATGTTAGCTT	6876
QY	5905	CAGAAAGAGGTGGCACAATCTCTCCTCAGGATGAGAAATTTATTCACATTAATTTGAAG	5964	QY	6958	GAAATGGGAAGAAACCACTATGTAAATATTTAGACCAAGGAGTAAATTCGAAAGATTTT	7017
DB	5800	CAGAAAGAGGTGGCACAATCTCTCCTCAGGATGAGAAATTTATTCACATTAATTTGAAG	5859	DB	6877	GAAATGGGAAGAAACCACTATGTAAATATTTAGACCAAGGAGTAAATTCGAAAGATTTT	6936
QY	5965	CTGGATCAGCTCATCAGTGGCAGACCTGGAATCTGAAGCAGAGATTAACCGCG	6020	QY	7018	TAAATCCAGGGGCCAAAGTTACCCCTCATCTTCCGAATTTGAAATGTGCAACCTTAA	7077
DB	5860	CTGGATCAGCTCATCAGTGGCAGACCTGGAATCTGAAGCAGAGATTTAAGCCGGTTA	5919	DB	6937	TAAATCCAGGGGCCAAAGTTACCCCTCATCTTCCGAATTTGAAATGTGCAACCTTAA	6996
QY	6021	-----GCCAATAAGCAAGAAATTCCTTCTCTGCAACATGTTGAA	6057	QY	7078	GAAATATCTATGCTTCTCTCAGCTGCTT	7108
DB	5920	CTTTAGTTATAGAAAAATCCATCAAGCCCAATTAAGCAAGAAATCTCTCTGCAACATGTTGAA	5979	DB	6997	GAAATATCTATGCTTCTCTCAGCTGCTT	7027
QY	6058	GAGCTTTGCAAAACCAACCTTAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTT	6117	RESULT 4			
DB	5980	GAGCTTTGCAAAACCAACCTTAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTT	6039	ADH41614 standard; DNA; 6900 BP.			
QY	6118	CTTCAGGATCTTTCTCAACTGATGCTGCTGCTTGGAAATAGAGCAAAACCCGTTTC	6177	ADH41614;			
DB	6040	CTTCAGGATCTTTCTCAACTGATGCTGCTGCTTGGAAATAGAGCAAAACCCGTTTC	6099	XX			
QY	6178	CCAAACATAAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGACGTAGTGT	6237	XX			
DB	6100	CCAAACATAAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGACGTAGTGT	6156	XX			

Novel human nucleic acid NOV15b.

ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;

KW nootropic; antiparkinsonian; antiasthmatic; antiinfertility;
KW cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
KW AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX Homo sapiens.

OS WO2003102159-A2.
XX 11-DEC-2003.
XX 04-JUN-2003; 2003WO-US017573.
XX 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386845P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 07-JUN-2002; 2002US-0387083P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387856P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.
XX

PA (CURA-) CURAGEN CORP.
XX Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;
PI Billeman K, Eitenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigar M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;
XX
DR WPI; 2004-053467/05.
DR P-PSDB; ADH41615.
XX
PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy, or
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.
XX
PS Claim 20; SEQ ID NO 167; 1503pp; English.
XX
CC The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the nucleic acid sequence of the invention.
XX
SQ Sequence 6900 BP; 2323 A; 1365 C; 1297 G; 1915 T; 0 U; 0 Other;
Query Match 95.7%; Score 6799; DB 12; Length 6900;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 6858; Conservative 0; Mismatches 15; Indels 30; Gaps 2;
QY 106 ATGGATTTCCTTATCATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTGAT 165
DB 1 ATGGATTTCCTTATCATTTTCTTTTACCTTTTATTTGGGACTTCAGAGACACAGGTGAT 60
QY 166 GTTTCCAATGTCGTTCTGCTAGTACGATATACCATCTCTTCAATTTTACAACA 225
DB 61 GTTTCCAATGTCGTTCTGCTAGTACGATATACCATCTCTTCAATTTTACAACA 120
QY 226 TACACCTCACCTGTTTACTAGATAGTGACACCAATGTAAACAAACACAGGCTCCAGTC 285
DB 121 TACACCTCACCTGTTTACTAGATAGTGACACCAATGTAAACAAACACAGGCTCCAGTC 180
QY 286 TTCCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTGTCTTTGGAATACACCACT 345
DB 181 TTCCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTGTCTTTGGAATACACCACT 240
QY 346 AATCCAAATGGAGGATTTATATCTTACATTTGCAATATAGGAAGTTTGTCTGCTGATG 405
DB 241 AATCCAAATGGAGGATTTATATCTTACATTTGCAATATAGGAAGTTTGTCTGCTGATG 300
QY 406 CAAACAGTATATACCAAGTCAGATCAAAGCCAGACAGTCGTGAAGTTCTTCTTACTAAT 465
DB 301 CAAACAGTATATACCAAGTCAGATCAAAGCCAGACAGTCGTGAAGTTCTTCTTACTAAT 360
QY 466 CTTAATCTCGGAACCAATATGAAATTAAGTTGCTGCTGGAAGAAACAGTCTGCAATTGA 525
DB 361 CTTAATCTCGGAACCAATATGAAATTAAGTTGCTGCTGGAAGAAACAGTCTGCAATTGA 420
QY 526 GTGTTTAGTGATCCATTTCTCTTCCAACTGCAGAAAGTCTCCAGGAAAAGTGTGAAT 585
DB 421 GTGTTTAGTGATCCATTTCTCTTCCAACTGCAGAAAGTCTCCAGGAAAAGTGTGAAT 480

QY	586	CTCACAGTTGAGCCCTACACGCTTCAGCAGTTAAAGCTGATTTGGTATTTTACTTCGGCAA	645	1666	ATTACTGACATTTGCAGCTGAAACAGAGCTGTCTTATGTATTCAGGAGACTTGTACTTTCACT	1725
DB	481	CTCACAGTTGAGCCCTACACGCTTCAGCAGTTAAAGCTGATTTGGTATTTTACTTCGGCAA	540	1561	ATTACTGACATTTGCAGCTGAAACAGAGCTGTCTTATGTATTCAGGAGACTTGTACTTTCACT	1620
QY	646	CCAAATGGCAAAATTACACGCTTCAGAGTTAGTGTCAAAACATGCCAGAAAGTGGATAGTA	705	1726	GAGCACATGATTTAGTGTATCTGCTTTTCAACCATCATGGGAGAGGACCACCAACAGTTCTC	1785
DB	541	CCAAATGGCAAAATTACACGCTTCAGAGTTAGTGTCAAAACATGCCAGAAAGTGGATAGTA	600	1621	GAGCACATGATTTAGTGTATCTGCTTTTCAACCATCATGGGAGAGGACCACCAACAGTTCTC	1680
QY	706	GTGAAAGATGTCTCAATACAGAGTAGAGGACATTTTCACTGGGAAATTTGCCAGAAATCAAT	765	1786	AGTGTTAGGACAGCTCAGCAAGTGCAGAGCTCCATTTAAATTTATTAATTTATTAATATTT	1845
DB	601	GTGAAAGATGTCTCAATACAGAGTAGAGGACATTTTCACTGGGAAATTTGCCAGAAATCAAT	660	1681	AGTGTTAGGACAGCTCAGCAAGTGCAGAGCTCCATTTAAATTTATTAATTTATTAATATTT	1740
QY	766	GAGAAATAGTGAATCTTTTATTTAGAGTAGAGGACATTTTCACTGGGAAATTTGCCAGAAATCAAT	825	1846	AGTGTTCATCTATTTTGTATTTGGGATCTTCAGAAATATCCCAATGGAAAAATAACT	1905
DB	661	GAGAAATAGTGAATCTTTTATTTAGAGTAGAGGACATTTTCACTGGGAAATTTGCCAGAAATCAAT	720	1741	AGTGTTCATCTATTTTGTATTTGGGATCTTCAGAAATATCCCAATGGAAAAATAACT	1800
QY	826	ACACCTCCCATCGGCTACACACATTCATCAAGCACGTTGACACAGAAATGAGATCACTCT	885	1906	CACTATACGATTTATGCAATTTGGATACAAACAGAGCATTCAGATAACTACCATTA	1965
DB	721	ACACCTCCCATCGGCTACACACATTCATCAAGCACGTTGACACAGAAATGAGATCACTCT	780	1801	CACTATACGATTTATGCAATTTGGATACAAACAGAGCATTCAGATAACTACCATTA	1860
QY	886	GTGTGAAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTTATCAACATATCTT	945	1966	GATACAGCTTTTCTCATACAGGTTTAAAGAAATACAAATACAAATGAGAGTGGCA	2025
DB	781	GTGTGAAAGAGCCTATCAGTTTGTAGTGACACACTTGAGACCTTTATCAACATATCTT	840	1861	GATACAGCTTTTCTCATACAGGTTTAAAGAAATACAAATACAAATGAGAGTGGCA	1920
QY	946	TTTGAAGTTTCAGCTGCTCACTCACTGAAGCAGGTTATTTGATAGTAGTACGATTTGTCAAA	1005	2026	GCCTCAACCCACGATGGAGAAAGTCTTTGTCTGAGAAATATGACATCTTTGTGAGAACT	2085
DB	841	TTTGAAGTTTCAGCTGCTCACTCACTGAAGCAGGTTATTTGATAGTAGTACGATTTGTCAAA	900	1921	GCCTCAACCCACGATGGAGAAAGTCTTTGTCTGAGAAATATGACATCTTTGTGAGAACT	1980
QY	1006	CCAGAAATCAGTGCCTGAAGGACCAACCAAACTGCGTAAACAGGCAACATCAAGGAAAG	1065	2086	TCAGAGATGAAACCGGAATCATCACTCAAGATGTGGAAGTAAATTTGATTTTACCGCAGAT	2145
DB	901	CCAGAAATCAGTGCCTGAAGGACCAACCAAACTGCGTAAACAGGCAACATCAAGGAAAG	960	1981	TCAGAGATGAAACCGGAATCATCACTCAAGATGTGGAAGTAAATTTGATTTTACCGCAGAT	2040
QY	1066	TCCTTTTCAATTTATGGGACCCACCACTATAGTAAACAGGGAATTTTAGTTATAGAGTT	1125	2146	GAATAAGGTGGAAGTGCACACCCGAAAGCCCAATGGGATCATTTATTCCTTATGAA	2205
DB	961	TCCTTTTCAATTTATGGGACCCACCACTATAGTAAACAGGGAATTTTAGTTATAGAGTT	1020	2041	GAATAAGGTGGAAGTGCACACCCGAAAGCCCAATGGGATCATTTATTCCTTATGAA	2100
QY	1126	GAATATATGGACCATCAGGTGCGATTTTGGATACAGCACAAAGACCTCAAGTTTGA	1185	2206	GTGCTATATAAAAAATATAGATCTTTATATATGAAGAACACATCAACACAGACATAATA	2265
DB	1021	GAATATATGGACCATCAGGTGCGATTTTGGATACAGCACAAAGACCTCAAGTTTGA	1080	2101	GTGCTATATAAAAAATATAGATCTTTATATATGAAGAACACATCAACACAGACATAATA	2160
QY	1186	TTCACTAACTAACACATTTACATGTATGTCTATATTTGCGGCTGAACCAAGTGCA	1245	2266	TTAAGGAACCTTAAGACCTCACACCTCTATATAATTTCTGTAAAGGTCTTACACAGATTT	2325
DB	1081	TTCACTAACTAACACATTTACATGTATGTCTATATTTGCGGCTGAACCAAGTGCA	1140	2161	TTAAGGAACCTTAAGACCTCACACCTCTATATAATTTCTGTAAAGGTCTTACACAGATTT	2220
QY	1246	GGGACTGGGCCCAGTCAAATATTTCAATTTCACTCCACAGATGTTCCAGGGCAGTG	1305	2326	GGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAAGACTTCGGAGACTGTGCTGAT	2385
DB	1141	GGGACTGGGCCCAGTCAAATATTTCAATTTCACTCCACAGATGTTCCAGGGCAGTG	1200	2221	GGTCAATGGCAATCAGGTATCTTTTACTCTCTGTAAAGACTTCGGAGACTGTGCTGAT	2280
QY	1306	TTTGATTTTCACTTCAGAGGTAGAAATCCAGCAAGTAAAGTAACTTTGGAGAAACCA	1365	2386	AGTGCAACCAAGAAATATCACTTACAAAATATTTCTTCGGAGAGATTCAGCTATCATTT	2445
DB	1201	TTTGATTTTCACTTCAGAGGTAGAAATCCAGCAAGTAAAGTAACTTTGGAGAAACCA	1260	2281	AGTGCAACCAAGAAATATCACTTACAAAATATTTCTTCGGAGAGATTCAGCTATCATTT	2340
QY	1366	CGACAAACCAATGGAATTTATACCAATACCGAGTGAAGTGTAGTTCAGAGACAGGA	1425	2446	CTTCCCCCAAGTAGTCCCAGTCCGAATCATATAAAAAATATACAAATTTTATCTCAAGAGAAT	2505
DB	1261	CGACAAACCAATGGAATTTATACCAATACCGAGTGAAGTGTAGTTCAGAGACAGGA	1320	2341	CTTCCCCCAAGTAGTCCCAGTCCGAATCATATAAAAAATATACAAATTTTATCTCAAGAGAAT	2400
QY	1426	ATAATTTTGGAAATATCTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCATGGCT	1485	2506	AATGGAAATGAGGAAGAACTATAATAACAACTCTTTTAAACCAAACTTAAAGTACTG	2565
DB	1321	ATAATTTTGGAAATATCTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCATGGCT	1380	2401	AATGGAAATGAGGAAGAACTATAATAACAACTCTTTTAAACCAAACTTAAAGTACTG	2460
QY	1486	CCAGAAATTTGGAACATAGTAGAGCCCAATGGTAGGATTTATGAGGGTTCCAGCAGAGATG	1545	2566	AAGAATATACCAATATATCATTTGAGGTGTCTGTAGTACACTGAAAGGTGAAGAGTT	2625
DB	1381	CCAGAAATTTGGAACATAGTAGAGCCCAATGGTAGGATTTATGAGGGTTCCAGCAGAGATG	1440	2461	AAGAATATACCAATATATCATTTGAGGTGTCTGTAGTACACTGAAAGGTGAAGAGTT	2520
QY	1546	TCGCTGACCTTCACCTTCATTTTATATATAAAGCCATCCAGATAAAACTTTT	1605	2626	CGGAGTGTCCCATTAAGTATATCTGAGGAGGAAGTGTCTCTGATTTCTCCCCCTCAAGAC	2685
DB	1441	TCGCTGACCTTCACCTTCATTTTATATATAAAGCCATCCAGATAAAACTTTT	1500	2521	CGGAGTGTCCCATTAAGTATATCTGAGGAGGAAGTGTCTCTGATTTCTCCCCCTCAAGAC	2580
QY	1606	CCTGCAAGGAATAGAGCTGAAGACCAAGCTTCAACAGTTGTAACTCAAGGAATCAGTAT	1665	2686	TTCTCTGTAAAAACAGTTGTCTGCTCAGGTTGAAGTGTCTGATGGCAACCAACCTGGAG	2745
DB	1501	CCTGCAAGGAATAGAGCTGAAGACCAAGCTTCAACAGTTGTAACTCAAGGAATCAGTAT	1560	2581	TTCTCTGTAAAAACAGTTGTCTGCTCAGGTTGAAGTGTCTGATGGCAACCAACCTGGAG	2640
				2746	CCAAATGGAATTTATCTTTATTTACACAGTTTATGTCTGGAATAGATCATCATTTAAAAA	2805

Db	2641		CCAAATGGAATTAATCTCTTTATTAACACAGTTTATGTCTCTGGAAATAGATCATCATTTAAAAACT	2700
Qy	2806	ATTAATGTCTACCTGAAACATCATCTGGAGTTATCAGATTTGGATTGGAATATAATGTTGGAATACAGT	2865	
Db	2701	ATTAATGTCTACCTGAAACATCATCTGGAGTTATCAGATTTGGATTGGAATATAATGTTGGAATACAGT	2760	
Qy	2866	GCTTATGTAAACAGCTAGCACACAGATTTGGTGATGGGAAAAACAGGAAGCAATATCATTTAGC	2925	
Db	2761	GCTTATGTAAACAGCTAGCACACAGATTTGGTGATGGGAAAAACAGGAAGCAATATCATTTAGC	2820	
Qy	2926	TTTTCAAAACCCAGAGGGAGCACCAAGCGATCTCTCCAAAAGATGTTTATTTATGCAACCTC	2985	
Db	2821	TTTTCAAAACCCAGAGGGAGCACCAAGCGATCTCTCCAAAAGATGTTTATTTATGCAAACTC	2880	
Qy	2986	AGTTCTTTTCATCAATTAATCTTTTCTGGACACCTCTCTTCAAAAACCTAATGGATTTATACAA	3045	
Db	2881	AGTTCTTTTCATCAATTAATCTTTTCTGGACACCTCTCTTCAAAAACCTAATGGATTTATACAA	2940	
Qy	3046	TATTACTCTGTGTTTATTAACAGAAATACTTTCAGGTACTTTTATGACAGAAATTTTACATCCAT	3105	
Db	2941	TATTACTCTGTGTTTATTAACAGAAATACTTTCAGGTACTTTTATGACAGAAATTTTACATCCAT	3000	
Qy	3106	GAACCTAACCAATGACTTTTGACAATATGACTGTATGCCAATATTAAGATAAATCGACAATA	3165	
Db	3001	GAAGTAACCAATGACTTTTGACAATATGACTGTATGCCAATATTAAGATAAATCGACAATA	3060	
Qy	3166	TTACAGCTACTATATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGCGAATGGGAATAAAGC	3225	
Db	3061	TTACAGCTACTATATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGCGAATGGGAATAAAGC	3120	
Qy	3226	AGTGACATCAATGGAAGTATACACAGATCAACAGACATACCTCAAGGGTTGTTGTGAAACCTG	3285	
Db	3121	AGTGACATCAATGGAAGTATACACAGATCAACAGCGTCCCTGAGGGTTGTTGTGAAACCTG	3180	
Qy	3286	ACTTACGAATCCATTTTCGTCAACTGCAATTAATGTAAGCTGGGTCCCAACGGGTCAACCA	3345	
Db	3181	ACTTACGAATCCATTTTCGTCAACTGCAATTAATGTAAGCTGGGTCCCAACGGGTCAACCA	3240	
Qy	3346	AACGGTCTAGTCTTCTACTATGTTTCTACTGATCTTTACAGACAGACTCTCTGCCATGTGAGA	3405	
Db	3241	AACGGTCTAGTCTTCTACTATGTTTCTACTGATCTTTACAGACAGACTCTCTGCCATGTGAGA	3300	
Qy	3406	CCACTCTTGTGTACATATGAGAGAAGCATATATTTGTGATAATCTGGAATAATACACTGAT	3465	
Db	3301	CCACTCTTGTGTACATATGAGAGAAGCATATATTTGTGATAATCTGGAATAATACACTGAT	3360	
Qy	3466	TATATATTAATAAATTAATCTCCATCAACAGAAAAGGGATTTCTCTGATACCTATATCTGCCAG	3525	
Db	3361	TATATATTAATAAATTAATCTCCATCAACAGAAAAGGGATTTCTCTGATACCTATATCTGCCAG	3420	
Qy	3526	CTATACATCAAGACTGGAAGAGATGTCCTCCAGAACTTCAACCAATATCAACACTTTTAAA	3585	
Db	3421	CTATACATCAAGACTGGAAGAGATGTCCTCCAGAAACTTTCACCAATAATCAACACTTTTAAA	3480	
Qy	3586	AACCTTTCTCTACTCTCAGTTCTCTTATCATGGATCCCCAGTAAAGCCAAATGGTGCA	3645	
Db	3481	AACCTTTCTCTACTCTCAGTTCTCTTATCATGGATCCCCAGTAAAGCCAAATGGTGCA	3540	
Qy	3646	ATAATAAGTTATGATTTTAACCTTTTACAAAGGCCAAATGAAAATTAATCTTTTCAATTTCT	3705	
Db	3541	ATAATAAGTTATGATTTTAACCTTTTACAAAGGCCAAATGAAAATTAATCTTTTCAATTTCT	3600	
Qy	3706	GATTAATTTACATAATATTTGGAAGAGCTTTTACCATTATACATATATAGCTTTTTTGTGCC	3765	
Db	3601	GATAATTTACATAATATTTGGAAGAGCTTTTACCATTATATAGCTTTTTTGTGCC	3660	
Qy	3766	GCAAGAACTAGAAAAGGACTTTGGTCTTCTCAGTATCTTTTTTCTTTTACACAGATGAGTCA	3825	
Db	3661	GCAAGAACTAGAAAAGGACTTTGGTCTTCTCAGTATCTTTTTTCTTTTACACAGATGAGTCA	3720	
Qy	3826	GTGCCGTTAGCACCTCCACAAAAATTTGCACTTTAATCAACTGTACTTTCAGACTTTGTATGG	3885	

4966	Qy	GCTGCATATGTTAGAAGGAAAGTCAAAGTGCCTGAAATGATTTGTTACTACTTTAGAAATCAGCC	5025
4861	Db	GCTGCATATGTTAGAAGGAAAGTCAAAGTGCCTGAAATGATTTGTTACTACTTTAGAAATCAGCC	4920
5026	Qy	CCAAAGGACCCCACTTAAACAATGACATTTTCAGAGAATACCCAGATGAAGTTACAAAAATTT	5085
4921	Db	CCAAAGGACCCCACTTAAACAATGACATTTTCAGAGAATACCCAGATGAAGTTACAAAAATTT	4980
5086	Qy	CAATTAACGTTTCCTTCCCTCTCTCAACCTTAATGGAATAATCCAAGTATATCAAGCTCTCG	5145
4981	Db	CAATTAACGTTTCCTTCCCTCTCTCAACCTTAATGGAATAATCCAAGTATATCAAGCTCTCG	5040
5146	Qy	GTTTACCGGAAGAATGATCTCTGCTGTCAGATTCACAACTCTCAGTATTTATACAGAAA	5205
5041	Db	GTTTACCGGAAGAATGATCTCTGCTGTCAGATTCACAACTCTCAGTATTTATACAGAAA	5100
5206	Qy	ACCAACACATTCGTCAATTCGAATCTGTAGAAGGACTAAAGGTGGAATACATACATCAATATC	5265
5101	Db	ACCAACACATTCGTCAATTCGAATCTGTAGAAGGACTAAAGGTGGAATACATACATCAATATC	5160
5266	Qy	AGTGTCTTACGCAGTCAATAGTGTCTGTCAGGTCCAAAGGTTCCGATGAGAAATAACCATG	5325
5161	Db	AGTGTCTTACGCAGTCAATAGTGTCTGTCAGGTCCAAAGGTTCCGATGAGAAATAACCATG	5220
5326	Qy	GATATCAAAGCTCCAGCACGACCAACCAAAACCAAAACCCCTATTTATGATGCCACAGGA	5385
5221	Db	GATATCAAAGCTCCAGCACGACCAACCAAAACCCCTATTTATGATGCCACAGGA	5280
5386	Qy	AAACTGCTTTGTGACTTCAACAAACATTTACAAATCAGAATGCCAATATGTTACTACAGTGAT	5445
5281	Db	AAACTGCTTTGTGACTTCAACAAACATTTACAAATCAGAATGCCAATATGTTACTACAGTGAT	5340
5446	Qy	GATCATGGACCAATAAAAAATGTACAAGTGTCTTCGACAGAAAACAGAGCTCAGCATGAT	5505
5341	Db	GATCATGGACCAATAAAAAATGTACAAGTGTCTTCGACAGAAAACAGAGCTCAGCATGAT	5400
5506	Qy	GGAAATGTAAACAAAGTGGTATGATGCATATTTTAAATAAGCAAGGCCATATTTTACAAT	5565
5401	Db	GGAAATGTAAACAAAGTGGTATGATGCATATTTTAAATAAGCAAGGCCATATTTTACAAT	5460
5566	Qy	GAAGGCTTTCTTAACCTCCATGTACAGAGAAGGAAGACAAAGTTTACTGGCAATGAAGAA	5625
5461	Db	GAAGGCTTTCTTAACCTCCATGTACAGAGAAGGAAGACAAAGTTTACTGGCAATGAAGAA	5520
5626	Qy	ATCTACATCATPAGTGTGTAATGTCATGATATTCCTGGCAATGAAGACAAAATTTGC	5685
5521	Db	ATCTACATCATPAGTGTGTAATGTCATGATATTCCTGGCAATGAAGACAAAATTTGC	5580
5686	Qy	AATGGACCATGGAACCAAAAGCAATACCTTATTTAAATTTAGAGCTACAAATATATG	5745
5581	Db	AATGGACCATGGAACCAAAAGCAATACCTTATTTAAATTTAGAGCTACAAATATATG	5640
5746	Qy	GGACAAATTTACTGACTCTGATTTATCTGACCCTGTGAAGACTTTTAGGGGAAGGACTTTCA	5805
5641	Db	GGACAAATTTACTGACTCTGATTTATCTGACCCTGTGAAGACTTTTAGGGGAAGGACTTTCA	5700
5806	Qy	GAAGAAGCCGTAGAGATCATTTCTTCGGTCACTTTGTGTATCCTTTTCAATTAATCTCCTT	5865
5701	Db	GAAGAAGCCGTAGAGATCATTTCTTCGGTCACTTTGTGTATCCTTTTCAATTAATCTCCTT	5760
5866	Qy	GGAAACAGCTATTTTTTGCATTTGCAAGAAATTCGACAGAAAGCAAGAGGTGGCAATAC	5925
5761	Db	GGAAACAGCTATTTTTTGCATTTGCAAGAAATTCGACAGAAAGCAAGAGGTGGCAATAC	5820
5926	Qy	TCTCTCTCAGGATGCAGAAATTTATGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTG	5985
5821	Db	TCTCTCTCAGGATGCAGAAATTTATGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTG	5880
5986	Qy	GCAGACCTGGAACTTGAAGGACGAGAGATTTAACGG-----	6020
5881	Db	GCAGACCTGGAACTTGAAGGACGAGAGATTTAACGGCGTTACTTAGTTATAGAAATCCATC	5940

RESULT 5
ABT06283
ID ABT
XX

Qy	6021	--GCCAATAAGCAAGAAATCCCTCTCTGCAACATGTTTGAAGAGCTTTTGCACAAAACAAC	6078
Db	5941	AAGCCAATAAGCAAGAAATCCCTCTCTGCAACATGTTTGAAGAGCTTTTGCACAAAACAAC	6000
Qy	6079	CTAAAGTTTCAAGAGAAATTTTCGGAAATACCAAAATTTCTTCAGGATCTTTCTTCAACT	6138
Db	6001	CTAAAGTTTCAAGAGAAATTTTCGGAAATACCAAAATTTCTTCAGGATCTTTCTTCAACT	6060
Qy	6139	GATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAACATATAAACCATATAAT	6198
Db	6061	GATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAACATATAAACCATATAAT	6117
Qy	6199	ATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCACGGTTCGGATTATATTAAT	6258
Db	6118	ATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCACGGTTCGGATTATATTAAT	6177
Qy	6259	GCCAGCTATATTTCTGCTGTTATTTATGTCCAAATGAATTTATGTCTACTCAAGGTCACACTA	6318
Db	6178	GCCAGCTATATTTCTGCTGTTATTTATGTCCAAATGAATTTATGTCTACTCAAGGTCACACTA	6237
Qy	6319	CCAGGAA CAGTTGAGATTTTTTGGAGAAATGGTGTGGGAAACCAAGGGCAAAAAATTAGTA	6378
Db	6238	CCAGGAA CAGTTGAGATTTTTTGGAGAAATGGTGTGGGAAACCAAGGGCAAAAAATTAGTA	6297
Qy	6379	ATGCTAACACAGTGTTTTGAAAAAGGAGCGATCAGATGCCATCAGTATTTTGGCCAGAGGAC	6438
Db	6298	ATGCTAACACAGTGTTTTGAAAAAGGAGCGATCAGATGCCATCAGTATTTTGGCCAGAGGAC	6357
Qy	6439	AACAAGCCAGTTACTGTCTTTGGAGATATAGTGAATACAAAGCTAATGGAAGGATGTTCAA	6498
Db	6358	AACAAGCCAGTTACTGTCTTTGGAGATATAGTGAATACAAAGCTAATGGAAGGATGTTCAA	6417
Qy	6499	ATAGATTGCACTATCAGGGATCTGAAATTTGAAAGGATCGGGANTTGCATGACTGTTCGA	6558
Db	6418	ATAGATTGCACTATCAGGGATCTGAAATTTGAAAGGATCGGGANTTGCATGACTGTTCGA	6477
Qy	6559	CAGTGTAACTTTTACTGCTCGCCAGACATGGGGTTCCTCAGAACACGCGCCCTCTAATT	6618
Db	6478	CAGTGTAACTTTTACTGCTCGCCAGACATGGGGTTCCTCAGAACACGCGCCCTCTAATT	6537
Qy	6619	CACTTTGTGAAGTTGGTTCGAGCAAGCAGGGCAATGACACCACTATGATGTTGTTCAAC	6678
Db	6538	CACTTTGTGAAGTTGGTTCGAGCAAGCAGGGCAATGACACCACTATGATGTTGTTCAAC	6597
Qy	6679	TGCAGTCTCGAGTTTGGAGAACTGGAGTTTTTATTTGCTCTGGACCACTTTAAACACACAT	6738
Db	6598	TGTAGTCTCGAGTTTGGAGAACTGGAGTTTTTATTTGCTCTGGACCACTTTAAACACACAT	6657
Qy	6739	ATAAATGACCACTGATTTTGTGGATATATATGGAAGTCTAGCTGAACTGAGAGAGTGAAGA	6798
Db	6658	ATAAATGACCACTGATTTTGTGGATATATATGGAAGTCTAGCTGAACTGAGAGAGTGAAGA	6717
Qy	6799	ATGTGCATGTGCAAGATCTGGCA CAGTATATCTTTTTTACACAGTGCATTCGGATCTC	6858
Db	6718	ATGTGCATGTGCAAGATCTGGCA CAGTATATCTTTTTTACACAGTGCATTCGGATCTC	6777
Qy	6859	TTATCAAAATGAGGAAGTAAATCAGCCCATCTGTTTTCGTAACCTTACAGCACTTCAGAAG	6918
Db	6778	TTATCAAAATGAGGAAGTAAATCAGCCCATCTGTTTTCGTAACCTTACAGCACTTCAGAAG	6837
Qy	6919	ATGGACTCTTTTGGACGGCATCGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCACTATG	6978
Db	6838	ATGGACTCTTTTGGACGGCATCGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCACTATG	6897
Qy	6979	TAA 6981	
Db	6898	TAA 6900	

RESULT 5
ABT06283
ID ABT06283 standard; cDNA; 6903 BP.
XX

QY	1360	AAACACGCAACCAAAATGGAATTAATAACCAATACCGAGTGAAAGTCTAGTTCACGAG	1419		2338	GAGCTATCATCTCTCCCGCAAGTAGTCCCAATGGAATCATACAAAAATATACAAATTTAT	2397		
Db	1261	AAACACGCAACCAAAATGGAATTAATAACCAATACCGAGTGAAAGTCTAGTTCACGAG	1320		QY	2494	CTCAGAGAAAGTAATGGAATGAGAAAGAACTATAAATACAACTCTTTTAAACCCAAAAC	2553	
QY	1420	ACAGGAATTAATTTGGAAATACTTTTGCTCACTGGAAATAATGAGATATATAATGACCCC	1479		Db	2398	CTCAAGAGAAAGTAATGGAATGAGAAAGAACTATAAATACAACTCTTTTAAACCCAAAAC	2457	
Db	1321	ACAGGAATTAATTTGGAAATACTTTTGCTCACTGGAAATAATGAG---ATAAATGACCCC	1377		QY	2554	ATTTAAAGTACTGAAGAAATATACCCAAATATATCATTTGAGGTCTCTGCTAGTAGTCACTGAAA	2613	
QY	1480	ATGGCTCCAGAAATCTGGAACATAGTAGAGCCCAATGGTAGGATATATGAGGGTTCAGCA	1539		Db	2458	ATT-----CTGAGAGAAATATACCCAAATATATCATTTGAGGTCTCTGCTAGTAGTCACTCAAA	2511	
Db	1378	ATGGCTCCAGAAATCTGGAACATAGTAGAGCCCAATGGTAGGATATATGAGGGTTCAGCA	1437		QY	2614	GGTGAAGGAGTTGCGAGTCTCCCAATAAGTATATCTGACGGAAGAAAGATGCTCCTGATTTCT	2673	
QY	1540	GAGATGCTGCTGACCTTCACTCACTTGCTACATTTATATAAACAAGCCATCCAGATAAA	1599		Db	2512	GGTGAAGGAGTTGCGAGTCTCCCAATAAGTATATCTGACGGAAGAAAGATGCTCCTGATTTCT	2571	
Db	1438	GAGATGCTGCTGACCTTCACTCACTTGCTACATTTATATAAACAAGCCATCCAGATAAA	1497		QY	2674	CCCCCTCAAGACTTCTGTATAAACAAGTTGCTGTGTGTCACGGTGAAGTTGTCATGGCAA	2733	
QY	1600	AACCTTTCTGCAAGGAATAGAGCTGAAGAACCAGACTTCAACAGTTGTAACTACAAGGAAT	1659		Db	2572	CCCCCTCAAGACTTCTGTATAAACAAGTTGCTGTGTGTCACGGTGAAGTTGTCATGGCAA	2631	
Db	1498	AACCTTTCTGCAAGGAATAGAGCTGAAGAACCAGACTTCAACAGTTGTAACTACAAGGAAT	1557		QY	2734	CCACCCCTGGAGCCAAATGGAATTAATCTTTTATTAACACAGTTTATATGTCT---GGAATAGA	2790	
QY	1660	CAGTATATTAATGACATTTGCGAGCTGAACAAGCTGTCTTATGTTATCAGGAGACTTGTACCT	1719		Db	2632	CCACCCCTGGAGCCAAATGGAATTAATCTTTTATTAACACAGTTTATATGTCTGGAGGAATAGA	2691	
Db	1558	CAGTATATTAATGACATTTGCGAGCTGAACAAGCTGTCTTATGTTATCAGGAGACTTGTACCT	1617		QY	2791	TCATCATTTAAAAAATAATTAATGTCCTGAAAACATCATTTGGAGTTATFCAGATTTGGATTTAT	2850	
QY	1720	TTCACTGAGCACAATGATTAATGATCTGCTTTTCAACCATCATGGGAGAAGGACCAACCAACA	1779		Db	2692	TCATCATTTAAAAAATAATTAATGTCCTGAAAACATCATTTGGAGTTATFCAGATTTGGATTTAT	2751	
Db	1618	TTCACTGAGCACAATGATTAATGATTAATGATCTGCTTTTCAACCATCATGGGAGAAGGACCAACCAACA	1677		QY	2851	AAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACACAGATTTGGTGTGGAAGAAACAGCA	2910	
QY	1780	GTTCTCAGTGTTAGGACAGCTCAGCAAGTGCCAAAGCTCCATTAATAAATTAATAAATAA	1839		Db	2752	AAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACACAGATTTGGTGTGGAAGAAACAGCA	2811	
Db	1678	GTTCTCAGTGTTAGGACAGCTCAGCAAGTGCCAAAGCTCCATTAATAAATTAATAAATAA	1737		QY	2911	AGCAATATCATTTAGCTTTTCAAAACACACAGAGGAGAGACCAAGCGATCTCTCCAAAGATGTT	2970	
QY	1840	AATATTAGTCTTCATCTATTTGTTTATATGGGATCTCCAGAAATATCCCAATGGAANA	1899		Db	2812	AGCAATATCATTTAGCTTTTCAAAACACACAGAGG---ACCAAGGAGATCTCTCCAAAGATGTT	2868	
Db	1738	AATATTAGTCTTCATCTATTTGTTTATATGGGATCTCCAGAAATATCCCAATGGAANA	1797		QY	2971	TATTATGCAAAACCTCAGTCTTTTCAAAACCTCAGTCTTTTCAAAACCTCAGTCTTTTCAAAACCT	3030	
QY	1900	ATAACTCATAACGATTAATGCAATGGAAATGGATACAAACAGAGCAATTCAGATAAAT	1959		Db	2869	TATTATGCAAAACCTCAGTCTTTTCAAAACCTCAGTCTTTTCAAAACCTCAGTCTTTTCAAAACCT	2928	
Db	1798	ATHACTCACTAAGCAATTTATGCAATGGAAATGGAAATGGAAATGGAAATGGAAATGGAAAT	1857		QY	3031	AAATGGAATTAACAATATTAATCTGTTTATTAACAGAAATCTTCAAGTACTTTTATGTCAG	3090	
QY	1960	ACCATAGATAACAGCTTTCTCAATAC-----AGGGTTAAAGAAATACACAAATAACAA	2013		Db	2929	AAATGGAATTAACAATATTAATCTGTTTATTAACAGAAATCTTCAAGTACTTTTATGTCAG	2988	
Db	1858	ACCATAGATAACAGCTTTCTCAATAC-----AGGGTTAAAGAAATACACAAATAACAA	1917		QY	3091	AAATTTTACACTCCATGAACTAACCAATGACTTTTGCAATATGACTGTATCCCAATATATA	3150	
QY	2014	ATGAGAGTGCCAGCTCAACCCAGATGGAGAAAGTCTTTGCTGCAAGAAATGACATC	2073		Db	2989	AAATTTTACACTCCATGAACTAACCAATGACTTTTGCAATATGACTGTATCCCAATATATA	3048	
Db	1918	ATGAGAGTGCCAGCTCAACCCAGTGGAGAAAGTCTTTGCTGCAAGAAATGACATC	1977		QY	3151	GATAAACTGCAATATTTAGCTACTATATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGGA	3210	
QY	2074	TTTGTGAGAACTCAGAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTAAATTTGAT	2133		Db	3049	GATAAACTGCAATATTTAGCTACTATATACATTTTGGTTTAAACAGCAAGTACTTTCAGTTGGA	3108	
Db	1978	TTTGTGAGAACTCAGAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTAAATTTGAT	2037		QY	3211	AAATGGAATTAAGGAGTGAACATCTTGAAGTATACACAGATCAAGACATCTGGAAGGG	3270	
QY	2134	GTTTACCGCAGATGAATAAGGTTGAAGTGTCAACCCGGAAGCCCAATGGGATCAAT	2193		Db	3109	AAATGGAATTAAGGAGTGAACATCTTGAAGTATACACAGATCAAGACCTCTCCGGAAGGG	3168	
Db	2038	GTTTACCGCAGATGAATAAGGTTGAAGTGTCAACCCGGAAGCCCAATGGGATCAAT	2097		QY	3271	TTTGTGGAAGAACTGACTTTACGAATCCATTTGCTCAACTGCAATAAATGTAAGCTGGGTC	3330	
QY	2194	ATTGCTTATGAAGTGTATATAAATAATAGATACATTTTATATATGAAGACATCAACA	2253		Db	3169	TTTGTGGAAGAACTGACTTTACGAATCCATTTGCTCAACTGCAATAAATGTAAGCTGGGTC	3228	
Db	2098	ATTGCTTATGAAGTGTATATAAATAATAGATACATTTTATATGAAGAACAATCAACA	2157		QY	3331	CCAACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCTCATGATCTTACAGCAGACT	3390	
QY	2254	ACAGACATATAATTAAGGAATTAAGACCTCAACCCCTCTATAAACAATTTCTGTAAGGTCT	2313		Db	3229	CCAACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCTCATGATCTTACAGCAGACT	3288	
Db	2158	ACAGACATATAATTAAGGAATTAAGACCTCAACCCCTCTATAAACAATTTCTGTAAGGTCT	2217		QY	3391	CCTCGCATGTGAGACCACTCTTGTATACATATGAGAGAGCATATTTTGTATATCTG	3450	
QY	2314	TACACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG	2373		Db	3289	CCTCGCATGTGAGACCACTCTTGTATACATATGAGAGAGCATATTTTGTATATCTG	3348	
Db	2218	TACACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG	2277		QY	3451	GAATAATACACTGATTATATATATAAATAATTAATCTCAACAGAAAGGATTTCTCTGAT	3510	
QY	2374	ACTGTGCTGATAGTGCACAGAAATATCACTTACAAAAATTTCTTCTGAGAGATT	2433		Db	3349	GAATAATACACTGATTATATATAAATAATTAATCTCAACAGAAAGGATTTCTCTGAT	3408	
Db	2278	ACTGTGCTGATAGTGCACAGAAATATCACTTACAAAAATTTCTTCTGAGAGATT	2337		QY	3511	ACCTATCTGCCCAGCTATATCATCAAGACTGGAAGAGATGTCCAGAAACTTCCACAAATA	3570	
QY	2434	GAGCTATCATCTCTCCCGCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTAT	2493						

QY 5731 GCTACAAATATTATGGGCAATTTACTGACTCTGATTATTTCTGACCTGTTAAGACTTTA 5790
DB 5629 GCTACAAATATTATGGGCAATTTACTGACTCTGATTATTTCTGACCTGTTAAGACTTTA 5688
QY 5791 GGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCAATTTTCGGTCACTTTTGTGATCTTT 5850
DB 5689 GCGGAAGGACTTTTCAGAAAGAACCGTAGAGATCAATTTTCGGTCACTTTTGTGATCTTT 5748
QY 5851 TCAATAATTTCTCTTGGACAGCTATTTTTCATTTGCAATTTGCAAGATTTGCAAGAGACAGAA 5910
DB 5749 TCAATAATTTCTCTTGGACAGCTATTTTTCATTTGCAAGATTTGCAAGAGACAGAA 5808
QY 5911 GAAGGTGGCACAATCTCTCTCAGGATGAGAAATTTATGACATAAATTTGAAGCTGGAT 5970
DB 5809 GAAGGTGGCACAATCTCTCTCAGGATGAGAAATTTATGACATAAATTTGAAGCTGGAT 5868
QY 5971 CAGCTCATCAGAGTGGCAGACTTGGAACTTGAAGGACGAGAGATTTAACGCG----- 6020
DB 5869 CAGCTCATCAGAGTGGCAGACTTGGAACTTGAAGGACGAGAGATTTAACGCGTTACTTACT 5928
QY 6021 -----GCCAATAGCAAGAAATCTCTCTGCAATGTTTGAAGGCTT 6063
DB 5929 TATAGAAATTCATCAAGCCAATAGCAAGAAATCTCTCTGCAATGTTTGAAGGCTT 5988
QY 6064 TGCACAAACAAACCTAAAGTTTCAAGAAAGAAATTTTCGAAATTTACCAAAATTTCTTCAG 6123
DB 5989 TGCACAAACAAACCTAAAGTTTCAAGAAAGAAATTTTCGAAATTTACCAAAATTTCTTCAG 6048
QY 6124 GATCTTTCTCACTGATCTGATCTGCTTGGATAGAGCAAAACCGTTTCCAAAC 6183
DB 6049 GATCTTTCTCACTGATCTGCTTGGATAGAGCAAAACCGTTTCCAAAC 6108
QY 6184 ATAAACCAATATAATTAATTAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGGT 6243
DB 6109 ATAAACCAATATAATTAATTAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGGT 6165
QY 6244 TCGGATTATTAATTAATTCAGCTGATATTTCTGCTGTTATTTATGTCCTCAATGAAATTTATGCT 6303
DB 6166 TCGGATTATTAATTAATTCAGCTGATATTTCTGCTGTTATTTATGTCCTCAATGAAATTTATGCT 6225
QY 6304 ACTCAAGGTCCTACTACCAAGGACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACCCAGG 6363
DB 6226 ACTCAAGGTCCTACTACCAAGGACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACCCAGG 6285
QY 6364 GCATAAATATTAGTAATCTAACACAGTGTGTTTGAAGAGGAGGATCAGATGCCATCAG 6423
DB 6286 GCATAAATATTAGTAATCTAACACAGTGTGTTTGAAGAGGAGGATCAGATGCCATCAG 6345
QY 6424 TATTGGCCAGAGGACAAACAGCAGTACTGCTTTGGAGATATAGTATTAACAAGCTA 6483
DB 6346 TATTGGCCAGAGGACAAACAGCAGTACTGCTTTGGAGATATAGTATTAACAAGCTA 6405
QY 6484 ATGGAGGATGTTCAATATAGATGGATCTACGGATCTGAAATTTGAAAGGATGGGAT 6543
DB 6406 ATGGAGGATGTTCAATATAGATGGATCTACGGATCTGAAATTTGAAAGGATGGGAT 6465
QY 6544 TGCATGACTGTTGCAAGTGTAACTTTACTGCTGCGCAGAGATGGGTTCTGAGAC 6603
DB 6466 TGCATGACTGTTGCAAGTGTAACTTTACTGCTGCGCAGAGATGGGTTCTGAGAC 6525
QY 6604 AGCGCCCTCTAATTCATTTTGAAGTTGGTTCGAGCAAGGAGGACATGACACCA 6663
DB 6526 AGCGCCCTCTAATTCATTTTGAAGTTGGTTCGAGCAAGGAGGACATGACACCA 6585
QY 6664 CCTATGATGTTCACTGCACTGCTGGAGTTGGAAGAACTGGAGTTTATTTGCTCTGGAC 6723
DB 6586 CCTATGATGTTCACTGCACTGCTGGAGTTGGAAGAACTGGAGTTTATTTGCTCTGGAC 6645
QY 6724 CATTTAAACAAATATAATGACCAATGATTTTGTGATATATATGCACTAGTAGTAA 6783
DB 6646 CATTTAAACAAATATAATGACCAATGATTTTGTGATATATATGCACTAGTAGTAA 6705
QY 6784 CTGAGAAGTGAAAGATGTGATGGTGCAGATCTGGCAGTATATCTTTTACACCAG 6843

DB 6706 CTGAGAAGTGAAAGAAATGTGCATGTTGGCAGAAATCTGGCAGATATCTTTTACACCAG 6765
QY 6844 TGCATTTCTGGATCTCTTATCAATAAGGAAGTAATCAGCCCATCTGTTTCTTAACATAT 6903
DB 6766 TGCATTTCTGGATCTCTTATCAATAAGGAAGTAATCAGCCCATCTGTTTCTTAACATAT 6825
QY 6904 TCAGCACTTCAGAGATGGACTCTTTGGACGCCCATGGAAGGTGATGTTGAGCTTGAATGG 6963
DB 6826 TCAGCACTTCAGAGATGGACTCTTTGGACGCCCATGGAAGGTGATGTTGAGCTTGAATGG 6885
QY 6964 GAAGAAACCACTATGTAA 6981
DB 6886 GAAGAAACCACTATGTAA 6903
RESULT 6
ADH41624
ID ADH41624 standard; DNA; 6903 BP.
XX AC ADH41624;
XX DT 25-MAR-2004 (first entry)
XX Novel human nucleic acid NOV15g.
DE ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;
KW anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;
KW nootropic; antiparkinsonian; antidiabetic; antinfertility;
KW cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
KW AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX Homo sapiens.
XX WO2003102159-A2.
XX 11-DEC-2003.
XX 04-JUN-2003; 2003WO-US017573.
XX 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386864P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 07-JUN-2002; 2002US-0387083P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387669P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.

PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.

PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
PI Burgess CE, Casman SU, Catterton E, Dhanabal M, Edinger SR;
PI Ellerman K, Eitenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
PI Khrantsov NV, Larochelle WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rotherberg ME, Shenoy SG, Shimkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernhet CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;

WPI; 2004-053467/05.
P-PSDB; ADH41625.

PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.

PS Claim 20; SEQ ID NO 177; 1503pp; English.

CC The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the nucleic acid sequence of the invention.

SQ Sequence 6903 BP; 2320 A; 1369 C; 1300 G; 1914 T; 0 U; 0 Other;

Query Match 94.0%; Score 6683.6; DB 12; Length 6903;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 6842; Conservative 0; Mismatches 19; Indels 57; Gaps 9;

QY 106 ATGGATTTCTTATCATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGAT 165
DB 1 ATGGATTTTCTTATCATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGAT 60

QY 166 GTTTCCAATGTCCTTCTCTGCTACTAGTAGATATAACCAATCTCTTCAATTTCTTACAACA 225
DB 61 GTTTCCAATGTCCTTCTCTGCTACTAGTAGATATAACCAATCTCTTCAATTTCTTACAACA 120
QY 226 TACACCTCACCTGCTTACTAGATAGTACACCAAAATGTACAAAACAGAGGCTCCAGTGC 285
DB 121 TACACCTCACCTGCTTACTAGATAGTACACCAAAATGTACAAAACAGAGGCTCCAGTGC 180
QY 286 TTCTAGCCGGGGAAGAGTCCGATCTGCTGGGATCTTCTGCTTTGGAAATACACCACT 345
DB 181 TTCTAGCCGGGGAAGAGTCCGATCTGCTGGGATCTTCTGCTTTGGAAATACACCACT 240
QY 346 AATCCAAATGGAAGGATATATCTTACATTTGTCCTCAATATAAGGAAGTTTCTCGTGGATG 405
DB 241 AATCCAAATGGAAGGATATATCTTACATTTGTCCTCAATATAAGGAAGTTTCTCGTGGATG 300
QY 406 CAAACAGTATATACCAAGTCCAGATCAAAAGCCAGAGTCTGGAAGTCTTCTTACTTAAT 465
DB 301 CAAACAGTATATACCAAGTCCAGATCAAAAGCCAGAGTCTGGAAGTCTTCTTACTTAAT 360
QY 466 CTTAATCTCTGGAAACAATATGAAATTAAGGTTGCTGCTGAAAACAGTCTGGCAATGGA 525
DB 361 CTTAATCTCTGGAAACAATATGAAATTAAGGTTGCTGCTGAAAACAGTCTGGCAATGGA 420
QY 526 GTCTTTAGTATGATCCTTTCTTTCCAAACCTGCAAGAGT--GCTCCAGGAAAGTGGTG 582
DB 421 GTCTTTAGTATGATCCTTTCTTTCCAAACCTGCAAGAGTCCAGTCTCCAGGAAAGTGGTG 480
QY 583 AATCTCACAGTTGAGGCTTACACGCTTCAAGCTTAAGCTGATTTGTTATTTACCTCGG 642
DB 481 AATCTCACAGTTGAGGCTTACACGCTTCAAGCTTAAGCTGATTTGTTATTTACCTCGG 540
QY 643 CAACCAATGGGAAAATTTACCAGCTTCAAGATTTAGTGTCAAAATGCTGCAAGTGGGATA 702
DB 541 CAACCAATGGGAAAATTTACCAGCTTCAAGATTTAGTGTCAAGATGCTGCAAGTGGGATA 600
QY 703 GTAGTCAAGATGTTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATGTCAGAGTGC 762
DB 601 GTAGTCAAGATGTTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATGTCAGAGTGC 660
QY 763 AAT---GAGAAATGATGAATCTTTTATGAGGTACAGCCAGCCCTTCTCCAACTTTGGT 819
DB 661 AATGTTAGAAATAGTGAATCTTTTATGAGGTACAGCCAGCCCTTCTCCAACTTTGGT 720
QY 820 AGAGTTTACACCTCCATCGGCTACCAACATTTCAATCAAGCACGTTGACACAGAAATGATC 879
DB 721 AGAGTTTACACCTCCATCGGCTACCAACATTTCAATCAAGCACGTTGACACAGAAATGATC 780
QY 880 AGCTCTGTGTGGAAGAGCTTATCAGTTTTGTAGTGACACACTTTGAGACCTTATACAACA 939
DB 781 AGCTCTGTGTGGAAGAGCTTATCAGTTTTGTAGTGACACACTTTGAGACCTTATACAACA 840
QY 940 TATCTTTTGAAGTTTCACTGCTACACCTGAGGAGGTTTATTTAGTAGTAGATGTC 999
DB 841 TATCTTTTGAAGTTTCACTGCTACACCTGAGGAGGTTTATTTAGTAGTAGATGTC 900
QY 1000 AGAACACCAAGATCAGTGTCTGAAGGACCAACCAAACTGCGTAAACAGGCAACATCACA 1059
DB 901 AGAACACCAAGATCAGTGTCTGAAGGACCAACCAAACTGCGTAAACAGGCAACATCACA 960
QY 1060 GGAAAGTCTCTTTTCAATTTTATGAGGACCCCACTATAGTAAACAGGAAATTTAGTTAT 1119
DB 961 GGAAAGTCTCTTTTCAATTTTATGAGGACCCCACTATAGTAAACAGGAAATTTAGTTAT 1020
QY 1120 AGAGTTGAATTTATATGAGGACCTCAGTCCGATTTTGGATTAACGACAAAGACCTCAAG 1179
DB 1021 AGAGTTGAATTTATGAGGACCTCAGTCCGATTTTGGATTAACGACAAAGACCTCAAG 1080
QY 1180 TTGTGATTCACATAACCTAACACCTTTTACAATGTATGATGTCTATATGCGGCTGAAC 1239
DB 1081 TTGTGATTCACATAACCTAACACCTTTTACAATGTATGATGTCTATATGCGGCTGAAC 1140

1240 AGTCAGGAGCTGGGCCCAAGTCAAAATATTTTCAGTATTCACCTCCACAGATGTTCCAGGG 1299
Db
1141 AGTCAGGAGCTGGGCCCAAGTCAAAATATTTTCAGTATTCACCTCCACAGATGTTCCAGGG 1200
Qy
1300 GCAGTGTGTTGATTTTCAAACTTGAGAGGTAGAAATCACGCAAGTAAAGATTTCTGGAAG 1359
Db
1201 GCAGTGTGTTGATTTTCAAACTTGAGAGGTAGAAATCACGCAAGTAAAGATTTCTGGAAG 1260
Qy
1360 AAACACGCAACCAATGGAATTTAACCATACCGAGTGAAGTGTAGTTCAGAG 1419
Db
1261 AAACACGCAACCAATGGAATTTAACCATACCGAGTGAAGTGTAGTTCAGAG 1320
Qy
1420 ACAGGAATATTTTGGAAATATCTTTGCTCAGTGAATTAATGAGTATTAATGAGTACCC 1479
Db
1321 ACAGGAATATTTTGGAAATATCTTTGCTCAGTGAATTAATGAGTATTAATGAGTACCC 1377
Qy
1480 ATGGCTCCAGAAATTTGGAACATAGTAGAGCCCAATGTTAGGATTTATATGAGGTTTCAGCA 1539
Db
1378 ATGGCTCCAGAAATTTGGAACATAGTAGAGCCCAATGTTAGGATTTATATGAGGTTTCAGCA 1437
Qy
1540 GAGATGTGTTGACCTTCACTCAGTGTCTACATTTTATATTAACAGCCATCCAGATAA 1599
Db
1438 GAGATGTGTTGACCTTCACTCAGTGTCTACATTTTATATTAACAGCCATCCAGATAA 1497
Qy
1600 AACTTTCTGCAAGGAATAGAGCTGAAGACCAAGTTCACCACTTGTAACTACAAGGAAT 1659
Db
1498 AACTTTCTGCAAGGAATAGAGCTGAAGACCAAGTTCACCACTTGTAACTACAAGGAAT 1557
Qy
1660 CAGTATATTACTGACATTTGAGCTGAACAGCTGCTTATGTTATACAGGACCTTGACCT 1719
Db
1558 CAGTATATTACTGACATTTGAGCTGAACAGCTGCTTATGTTATACAGGACCTTGACCT 1617
Qy
1720 TTCACTGAGCACATGATGATGCTGCTTTTCACTCATGAGAGGACCAACCA 1779
Db
1618 TTCACTGAGCACATGATGATGCTGCTTTTCACTCATGAGAGGACCAACCA 1677
Qy
1780 GTTCTCAGTGTAGGACAGCTGAGCAAGTCCCAAGCTCCATTAATAATAATAATAA 1839
Db
1678 GTTCTCAGTGTAGGACAGCTGAGCAAGTCCCAAGCTCCATTAATAATAATAATAA 1737
Qy
1840 AATPATTAGTCTTTCATCTAATTTTGTATATGTTGGATCTCCAGAAATATCCCAATGAAAA 1899
Db
1738 AATPATTAGTCTTTCATCTAATTTTGTATATGTTGGATCTCCAGAAATATCCCAATGAAAA 1797
Qy
1900 ATAACCTCATATACGATTTATGCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1959
Db
1798 ATAACCTCATATACGATTTATGCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1857
Qy
1960 ACCATAGATAACAGCTTTCTCATAAC - - - - - AGGGTTAAAGAAATACACAAATACAAA 2013
Db
1858 ACCATAGATAACAGCTTTCTCATACAGGTATAGGGTTAAAGAAATACACAAATACAAA 1917
Qy
2014 ATGAGAGTGGCAGCTTCAACCCACGATGGAGAAAGTCTTTGTCTGAAGAAAATGATC 2073
Db
1918 ATGAGAGTGGCAGCTTCAACCCACGTTGGAGAAAGTCTTTGTCTGAAGAAAATGATC 1977
Qy
2074 TTTGTGAGAACTTCAGAGATGAAACCGGAATCATCTCAAGATGCGAAGTAATTGAT 2133
Db
1978 TTTGTGAGAACTTCAGAGATGAAACCGGAATCATCTCAAGATGCGAAGTAATTGAT 2037
Qy
2134 GTTACCGCAGATGAATTAAGTTGAAAGTGTACACCCGAAAGCCCAATGGGATCAT 2193
Db
2038 GTTACCGCAGATGAATTAAGTTGAAAGTGTACACCCGAAAGCCCAATGGGATCAT 2097
Qy
2194 ATTTGCTTTATGAAGTGTATATAAATAATATAGATATCTTTATATATGAAGAACATCAACA 2253
Db
2098 ATTTGCTTTATGAAGTGTATATAAATAATATAGATATCTTTATATATGAAGAACATCAACA 2157
Qy
2254 ACAGACATATATATAGGAATTTAAGACCTCACACCTCTATTAACATTTCTGTAAGGTCT 2313
Db
2158 ACAGACATATATATAGGAATTTAAGACCTCACACCTCTATTAACATTTCTGTAAGGTCT 2217
Qy
2314 TACACACAGATTTGGTCTATGGCAATCAGGTATCTTTTACTCTCTGTAGGAGCTTCGGAG 2373

2218 TACACACAGATTTGGTCTATGGCAATCAGGTATCTTTTACTCTCTGTAGGAGCTTCGGAG 2277
Qy
2374 ACTGTGCTGTATAGTGGCCACGAAATATACCTTACAAATAATTTCTTCTCGAGAGAT 2433
Db
2278 ACTGTGCTGTATAGTGGCCACGAAATATACCTTACAAATAATTTCTTCTCGAGAGAT 2337
Qy
2434 GAGCTATCATTTCTTTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACAAATTTAT 2493
Db
2338 GAGCTATCATTTCTTTCCCAAGTAGTCCCAATGGAATCATATAAATAATATACAAATTTAT 2397
Qy
2494 CTCAAGAGAGTAATTTGGAAATCAGGAAAGAACTATAAATAAATCAACCTTTTAAACCCAAAAC 2553
Db
2398 CTCAAGAGAGTAATTTGGAAATCAGGAAAGAACTATAAATAAATCAACCTTTTAAACCCAAAAC 2457
Qy
2554 ATTAAGTAGTCAAGAAATATACCAATATATCATTTGAGGTCTGCTAGTACACATGAA 2613
Db
2458 ATT - - - - - CTGAAGAAATATACCAATATATCATTTGAGGTCTGCTAGTACACATGAA 2511
Qy
2614 GGTGAAGGAGTTTCGGAGTCTCCCAATAGTATACCTGACGAGGAAGATGCTCTGATTTCT 2673
Db
2512 GGTGAAGGAGTTTCGGAGTCTCCCAATAGTATACCTGACGAGGAAGATGCTCTGATTTCT 2571
Qy
2674 CCCCTCTCAAGACTTCTCTGTAAAAACAGTTGTCTGTGTCAACGGTGAAGTTGTCAATGGCAA 2733
Db
2572 CCCCTCTCAAGACTTCTCTGTAAAAACAGTTGTCTGTGTCAACGGTGAAGTTGTCAATGGCAA 2631
Qy
2734 CACCCCTGGAGCCAAATGGAATTTCTTTTATATACAGTTTATGTTCT - - - - - GGAATACA 2790
Db
2632 CACCCCTGGAGCCAAATGGAATTTCTTTTATATACAGTTTATGTTCTGGAAGGAATAGA 2691
Qy
2791 TCATCATTAATAAACTATTAAATGCTCACTGAAACATCATTTGGAGTTATCAGATTTGGATTTAT 2850
Db
2692 TCATCATTAATAAACTATTAAATGCTCACTGAAACATCATTTGGAGTTATCAGATTTGGATTTAT 2751
Qy
2851 AATGTTGAATAACAGTCTTTATGTAAACAGCTAGCAACAGATTTGGTGTATGGGAAAAACAGAA 2910
Db
2752 AATGTTGAATAACAGTCTTTATGTAAACAGCTAGCAACAGATTTGGTGTATGGGAAAAACAGAA 2811
Qy
2911 AGCAATATCATTTAGCTTTTCAACACAGAGGGGACCAAGCGATCTCCCAAGATGTT 2970
Db
2812 AGCAATATCATTTAGCTTTTCAACACAGAGGG - - - - - ACCAAGCGATCTCCCAAGATGTT 2868
Qy
2971 TATTATGCAAACTCAGTCTTTCATCAATAATTTCTTTCTGACACCTCTCTTCAAAACCT 3030
Db
2869 TATTATGCAAACTCAGTCTTTCATCAATAATTTCTTTCTGACACCTCTCTTCAAAACCT 2928
Qy
3031 AATGGGATTTATCAATTTACTCTGTTTATTAACAGAAATACTTTCAGGTAATTTTATGTCAG 3090
Db
2929 AATGGGATTTATCAATTTACTCTGTTTATTAACAGAAATACTTTCAGGTAATTTTATGTCAG 2988
Qy
3091 AATTTTACATCTCCATGAACTAACCAATGACTTTTGACAAATATGACTGTATCCCAATTTATA 3150
Db
2989 AATTTTACATCTCCATGAACTAACCAATGACTTTTGACAAATATGACTGTATCCCAATTTATA 3048
Qy
3151 GATAAACTCACAAATTTTCAGTCTATATACATTTTGGTTAAACAGCAAGTACTTTCAGTTGGA 3210
Db
3049 GATAAACTCACAAATTTTCAGTCTATATACATTTTGGTTAAACAGCAAGTACTTTCAGTTGGA 3108
Qy
3211 AATGGGAATAAAGCAGTGCATCATCATTTGAAATATACAGATCAAGACATACCTGAAGGG 3270
Db
3109 AATGGGAATAAAGCAGTGCATCATCATTTGAAATATACAGATCAAGACGCTCCCTGAAGGG 3168
Qy
3271 TTTGTGGAAAACCTGACTTACGAACTCATTTGTCGTCGCAATGCAATGCTGAGCTGGCTC 3330
Db
3169 TTTGTGGAAAACCTGACTTACGAACTCATTTGTCGTCGCAATGCAATGCTGAGCTGGCTC 3228
Qy
3331 CCACCGGCTCAACCAAAACGCTGAGTCTTCTACTATGTTTCTACTGATCTTTTACAGCAGACT 3390
Db
3229 CCACCGGCTCAACCAAAACGCTGAGTCTTCTACTATGTTTCTACTGATCTTTTACAGCAGACT 3288
Qy
3391 CCTCGCATGTGAGACCACTCTTGTGTTAATATGAGAGAGGCAATATATTTTGTATATCTG 3450

Db 3289 CCTCGCCTGAGACCACTCTTGTACATATGAGAGACATATATTTTGATATCTG 3348
Qy 3451 GAAAAATACACTGATATATATTTAAAAATTTACTCCATCAACAGAAAAGGATTTCTCTGAT 3510
Db 3349 GAAAAATACACTGATATATATTTAAAAATTTACTCCATCAACAGAAAAGGATTTCTCTGAT 3408
Qy 3511 ACCTATACCTGCCAGCTATACATCAAGCTGAGAGAGATGTCACAGAACTTCCACCAATA 3570
Db 3409 ACCTATACCTGCCAGCTATACATCAAGCTGAGAGAGATGTCACAGAACTTCCACCAATA 3468
Qy 3571 ATCAACACATTTTAAAAACCTTCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTA 3630
Db 3469 ATCAACACATTTTAAAAACCTTCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTA 3528
Qy 3631 AAGCCAAATGGTGCATATAAAGTTATGATTTTAACTTTTACAAGGACCAATGAAATTTAT 3690
Db 3529 AAGCCAAATGGTGCATATAAAGTTATGATTTTAACTTTTACAAGGACCAATGAAATTTAT 3588
Qy 3691 TCTTTTCATTTCTGATTAATACATTAATTTGGAAGAGCTTTTCAAGTTTACATTTAT 3750
Db 3589 TCTTTTCATTTCTGATTAATACATTAATTTGGAAGAGCTTTTCAAGTTTACATTTAT 3648
Qy 3751 AGCTTTTCTGCCGCAAGAACTAGAAAAGGACTTGGTCTTCCAGTATTTCTTTCTTT 3810
Db 3649 AGCTTTTCTGCCGCAAGAACTAGAAAAGGACTTGGTCTTCCAGTATTTCTTTCTTT 3708
Qy 3811 TACACAGATGAGTCAGTCGCGTTAGCACCTCCACAAAATTTGACTTTTAACTCACTGACT 3870
Db 3709 TACACAGATGAGTCAGTCGCGTTAGCACCTCCACAAAATTTGACTTTTAACTCACTGACT 3768
Qy 3871 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATTTGTTAAAGTA 3930
Db 3769 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATTTGTTAAAGTA 3828
Qy 3931 TATAGTTTAAATTCATGAACATGAATGAACCTGACACTATATATTAAGATATATCAGGA 3990
Db 3829 TATAGTTTAAATTCATGAACATGAATGAACCTGACACTATATTAAGATATATCAGGA 3888
Qy 3991 TTTTAAACTGAAGCCAAACTTGTGGACTGGAAACCACTGAGCACCTACTCTATCCGTGTA 4050
Db 3889 TTTTAAACTGAAGCCAAACTTGTGGACTGGAAACCACTGAGCACCTACTCTATCCGTGTA 3948
Qy 4051 TCTGCGTTCCAAAAGTTGGAATGGCAATCAATTTAGTAAATGATGATAAAATTCACAACC 4110
Db 3949 TCTGCGTTCCAAAAGTTGGAATGGCAATCAATTTAGTAAATGATGATAAAATTCACAACC 4008
Qy 4111 CAAGAAATCAAGTTCCAGATGTCGTCAGATATATCCAGTGCATGCGCACTAGCTGGCAGTCA 4170
Db 4009 CAAGAAATCAAGTTCCAGATGTCGTCAGATATATCCAGTGCATGCGCACTAGCTGGCAGTCA 4068
Qy 4171 GTTTTGTGAAATGGGATCCACCACAAAAGGCAAAATGGAAATATAACGCAATATATGGTA 4230
Db 4069 GTTTTGTGAAATGGGATCCACCACAAAAGGCAAAATGGAAATATAACGCAATATATGGTA 4128
Qy 4231 ACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGATCAATGATACACTTTTCAATAAG 4290
Db 4129 ACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGATCAATGATACACTTTTCAATAAG 4188
Qy 4291 CTTCTTGCCAAATACCTCATATGTTCTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGT 4350
Db 4189 CTTCTTGCCAAATACCTCATATGTTCTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGT 4248
Qy 4351 GATGAAAGCACATGCACTGTCAGCACACTACCTGAAACAGTTTCCAGTGTTCGCCAATAAT 4410
Db 4249 GATGAAAGCACATGCACTGTCAGCACACTACCTGAAACAGTTTCCAGTGTTCGCCAATAAT 4308
Qy 4411 ATTGCTTTTCTGATGTTTCACTGATGTCAGTGCACCAATTTGACATGGAATGACCTGACACT 4470
Db 4309 ATTGCTTTTCTGATGTTTCACTGATGTCAGTGCACCAATTTGACATGGAATGACCTGACACT 4368
Qy 4471 ATCCTTGGCTACTTTTCAAAAATTAACAAAATTTACCACTCAACTGCTGCTCAAAAATGCAAA 4530
Db 4369 ATCCTTGGCTACTTTTCAAAAATTAACAAAATTTACCACTCAACTGCTGCTCAAAAATGCAAA 4428

Qy 4531 GAATGGGAATCCGAAGAATGTGTGAAATATCAAAAAATTTCAATACCTCTATGAAAGTCTCAC 4590
Db 4429 GAATGGGAATCCGAAGAATGTGTGAAATATCAAAAAATTTCAATACCTCTATGAAAGTCTCAC 4488
Qy 4591 TTAACCTGAAGAGACAGTATATGGAATTTAAAGAAATTTTAGATGGTATAGATTTCCAAGTGGCT 4650
Db 4489 TTAACCTGAAGAGACAGTATATGGAATTTAAAGAAATTTTAGATGGTATAGATTTCCAAGTGGCT 4548
Qy 4651 GCCAGACCAATGCTGGCTATGGCAATGCTTCAAACTGGATTTTCTACAAAACTCTGCTCT 4710
Db 4549 GCCAGACCAATGCTGGCTATGGCAATGCTTCAAACTGGATTTTCTACAAAACTCTGCTCT 4608
Qy 4711 GGCCTCTCAGATGCTCTCTGAAATGTTTCAATGATAGTAGCAACATCACCTTTTAGCATC 4770
Db 4609 GGCCTCTCAGATGCTCTCTGAAATGTTTCAATGATAGTAGCAACATCACCTTTTAGCATC 4668
Qy 4771 AGCATAAAGCTGGAGTGAACCTGCTGTCATTTACTGGACCAACATGTTTCTGATGATGTC 4830
Db 4669 AGCATAAAGCTGGAGTGAACCTGCTGTCATTTACTGGACCAACATGTTTCTGATGATGTC 4728
Qy 4831 AAATCGGTAGATAATGATGAATTTTATATATCTTTCATCAAGTCAAAATGAAGAAATATAA 4890
Db 4729 AAATCGGTAGATAATGATGAATTTTATATATCTTTCATCAAGTCAAAATGAAGAAATATAA 4788
Qy 4891 ACCATAGAAATTAAGATTTAGAAATATTTCAAGGATATTTCTGATGATGATGATGATGAT 4950
Db 4789 ACCATAGAAATTAAGATTTAGAAATATTTCAAGGATATTTCTGATGATGATGATGATGAT 4848
Qy 4951 ACTGGGAACATTTAGTCTGCATATGTAAGGGAAGTCAAGTCTGAAATGATGATGATGATGAT 5010
Db 4849 ACTGGGAACATTTAGTCTGCATATGTAAGGGAAGTCAAGTCTGAAATGATGATGATGATGAT 4908
Qy 5011 ACTTTAGAAATCAAGCCAAAGGACCACTTAAACATGACATTTTCAAGAGATACACAGAT 5070
Db 4909 ACTTTAGAAATCAAGCCAAAGGACCACTTAAACATGACATTTTCAAGAGATACACAGAT 4968
Qy 5071 GAAGTTACAAAATTTCAATTAAGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5130
Db 4969 GAAGTTACAAAATTTCAATTAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5028
Qy 5131 GTATATCAAGCTCTGGTTTACCGAAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5190
Db 5029 GTATATCAAGCTCTGGTTTACCGAAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5088
Qy 5191 AGTATTATACAGAAAACCAACATTCGTCATTTGCAATGCTAGAAAGTCAATTAAGAGTGA 5250
Db 5089 AGTATTATACAGAAAACCAACATTCGTCATTTGCAATGCTAGAAAGTCAATTAAGAGTGA 5148
Qy 5251 CATACATACATATCAGTGTTTTACGCAATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5310
Db 5149 CATACATACATATCAGTGTTTTACGCAATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5208
Qy 5311 ATGAGAAATCAACATGATTAACAGCTCCAGACGACCAAAACCAAAACCAAAACCAAAACCAAA 5370
Db 5209 ATGAGAAATCAACATGATTAACAGCTCCAGACGACCAAAACCAAAACCAAAACCAAAACCA 5268
Qy 5371 TATGATGCCACAGGAAAACCTGCTTGTGACTTCAACAAATTTACAAATCAGAAATGCAAAATA 5430
Db 5269 TATGATGCCACAGGAAAACCTGCTTGTGACTTCAACAAATTTACAAATCAGAAATGCAAAATA 5328
Qy 5431 TGTACTACTACATGATCATGGAACCAATAAAAATGTAAGTGTCTGCGACAGAAACCA 5490
Db 5329 TGTACTACTACATGATCATGGAACCAATAAAAATGTAAGTGTCTGCGACAGAAACCA 5388
Qy 5491 GGAGCTCAGCATGATGGAATGTAACAAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 5550
Db 5389 GGAGCTCAGCATGATGGAATGTAACAAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 5448
Qy 5551 CCATATTTTACAAAATGAAGGCTTTCTTAACTTCCATGTCAGAAAGGAAAGCAAAAGTTT 5610
Db 5449 CCATATTTTACAAAATGAAGGCTTTCTTAACTTCCATGTCAGAAAGGAAAGCAAAAGTTT 5508

PT atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.
PS Claim 9; Page 23-25; 353pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC several novel human proteins, designated NOVX. These can be used in the
CC treatment of, amongst others, cancers, autoimmune diseases, infections,
CC inflammatory diseases, storage disorders, muscle disorders,
CC neurodegenerative diseases and developmental defects. The present
CC sequence is a coding sequence of the invention
XX
SQ Sequence 6994 BP; 2330 A; 1388 C; 1320 G; 1956 T; 0 U; 0 Other;
Query Match 90.8%; Score 6453; DB 6; Length 6994;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 6770; Conservative 0; Mismatches 215; Indels 48; Gaps 14;
QY 76 TGAATCTACTGGCTGAAATATGTAATAAGATGATTTCTTATCATTTTTCTTTACTT 135
DB 1 TGAATCTACTGGCTGAAATATGTAATAAGATGATTTCTTATCATTTTTCTTTACTT 60
QY 136 TTTATTGGGACTTCAGAGACACAGGTGTGATGTTTCCAAATGTCGTTCTCGTACTAGGTAC 195
DB 61 TTTATTGGGACTTCAGAGACACAGGTAGATGTTTCCAAATGTCGTTCTCGTACTAGGTAC 120
QY 196 GATATAACCATCTCTCAATTTCTAACAACACACCTCACCTGTTACTAGATAGTGACA 255
DB 121 GATATAACCATCTCTCAATTTCTAACAACACACCTCACCTGTTACTAGATAGGTGCT 180
QY 256 CCAATGTACAAACCCAGGGCTCCAGTCTTCTAGCGGGGAAAGAGTCGGATCTGCT 315
DB 181 TCTAATG-----AACAGGGCTCCAGTCTTCTAGCGGGGAAAGAGTCGGATCTGCT 234
QY 316 GGGATCTCTCTGTCTGGATATACACACCTTAATCCAAATGCGAAGGATTAATCTTACAT 375
DB 235 GGGATCTCTCTGTCTGGATATACACACCTTAATCCAAATGCGAAGGATTAATCTTACAT 294
QY 376 GTCAATATATAGGAAGTTTCTCGTGGATGCAACAGTATATACAAAGTCAGATCAAAG 435
DB 295 GTCAATATATAGGAAGTTTCTCGTGGATGCAACAGTATATACAAAGTCAGATCAAAG 354
QY 436 CCAGACAGTCTGGAAGTTCTTCTTACTTAATCTTAATCTTAACTCGGACCAACATATGAATTAAG 495
DB 355 CCAGACAGTCTGGAAGTTTCTTCTTACTTAATCTTAACTCGGACCAACATATGAATTAAG 414
QY 496 GTTGTGCTCAAAACAGTCTGCAATGGAGTGTGTTAGTGCATCTTCTCTTCCAAACT 555
DB 415 GTAGCTGCTGAAACAGTCTGCAATGGAGTGTGTTAGTGCATCTTCTTCCAAACT 474
QY 556 GCAGAAAGTCTCCAGGAAAGTGGTGAATCTCACAGTTGAGGCTCAACCGCTTCAGCA 615
DB 475 GCAGAAAGTCTCCAGGAAAGTGGTGAATCTCACAGTTGAGGCTGTC--CCGTTTCAGCA 532
QY 616 GTTAAGCTGATTTGTTATTTACTTCGCGACCAATGCGCAAAATACCAGCTTCAGATT 675
DB 533 G-TAAGCTGATGGTA--TACCCTGCGCAACCAAA-AAAAAATTTACAGCTTCAAGATT 588
QY 676 AGTGTCAACATGCCAGAGTGGATAGTAGTGAAGATGTCCTCAATCAGAGTAGAGGAC 735
DB 589 AGTGTCAAGCTAACAGAGTGGGATAGTAGTGAAGAGTGTCAATCAGAGTGGAGTGC 648
QY 736 ATTTTGACTGGGAAATTTGCC--AGAATGCAATGAGATAGTGAATCTTTTTTATGGAGT 792
DB 649 ATTTTGAAGTCTTCCCTTCTGCACTGCAACAGAGATAGTGAATCTTTTTTATGGAGT 708
QY 793 ACAGCAGCCCTCTCCAACTCTGTGATAGTATACCTTCATCGGCTACCAACATTCOA 852
DB 709 ACAGCAGCCCTCTCCAACTCTGTGATAGTATACCTTCATCGGCTACCAACATTCOA 768
QY 853 TCAAGCAGCTTGACACAGATGATGATCAGCTCTGTGTTGGAAAGAGCTATCAGTTTGTGA 912
DB 769 TCAAGCAGCTTGACACAGATGATGATCAGCTCTGTGTTGGAAAGAGCTATCAGTTTGTGA 825

QY 913 GTGACACACTTGAGACCTTATACACATATCTTTTTGAAAGTTTCAGCTGCTCAACCTGAA 972
DB 826 GTGACACACTTGAGACCTTATACACATATCTTTTTGAAAGTTTCAGCTGCTCAACCTGAA 885
QY 973 GCAGGTTATATGATAGTACGATTGTGAGAACACACAGAAATCAGTGTCTGAAGACCAACCA 1032
DB 886 GCAGGTTATATGATAGTACGATTGTGAGAACACACAGAAATCAGTGTCTGAAGACCAACCA 945
QY 1033 CAAACCTGCGTAAACAGGCAACATCAACAGGAAAGTCTTTTCAATTTTATGGGACCAACCA 1092
DB 946 CAAACCTGCGTAAACAGGCAACATCAACAGGAAAGTCTTTTCAATTTTATGGGACCAACCA 1005
QY 1093 ACTATAGTAAACAGGGAATTTAGTTATAGAGTTGAATATATGACCAT---CAGTCTGC 1149
DB 1006 ACTATAGTAAACAGGGAATTTAGTTATAGAGTTGAATATATGACCATCAGCAGGTCTGC 1065
QY 1150 ATTTTGGATAAACAGCAACAAAGACCTCAAGTTTGGCAATTCACAACTTAAACCATTTTACA 1209
DB 1066 ATTTTGGATAAACAGCAACAAAGACCTCAAGTTTGGCAATTCACAACTTAAACCATTTTACA 1125
QY 1210 ATGTATGATGTCTATATTCGGCTGAAACAGTGTGAGGAGCTGGGCCCAAGTCAATATATT 1269
DB 1126 ATGTATGATGTCTATATTCGGCTGAAACAGTGTGAGGAGCTGGGCCCAAGTCAATATATT 1185
QY 1270 TCAGTATTCCTCTCCACAGATGTTCCAGGGCAGTGTGTTGATTTTACAATTGACAGAGGTA 1329
DB 1186 TCAGTATTCCTCTCCACAGATGTTCCAGGGCAGTGTGTTGATTTTACAATTGACAGAGGTA 1245
QY 1330 GAATCCACCAAGTAGTAAGATTTCTTGGAAAGAACACAGCAACAAATGGAATTTATTAAC 1389
DB 1246 GAATCCACCAAGTAGTAAGATTTCTTGGAAAGAACACAGCAACAAATGGAATTTATTAAC 1305
QY 1390 CAATACCGAGTGAAGTGTCTAGTTCCAGAGACAGGAATTAATTTTGGAAATACTTTGCTC 1449
DB 1306 CAATACCGAGTGAAGTGTCTAGTTCCAGAGACAGGAATTAATTTTGGAAATACTTTGCTC 1365
QY 1450 ACTGGAATTAATGAGTATTAATGACCCCATGCTCCAGAAATTTGTGAACATGATAGAG 1509
DB 1366 ACTGGAATTAATGAG---ATAAATGACCCCATGCTCCAGAAATTTGTGAACATGATAGAG 1422
QY 1510 CCAATGGTAGGATTAATGAGGGTTCCAGCAGAGATGTCGCTGACCTTCACTCACTTGCT 1569
DB 1423 CCAATGGTAGGATTAATGAGGGTTCCAGCAGAGATGTCGCTGACCTTCACTCACTTGCT 1482
QY 1570 ACATTTTATATAACAGCCATCCAGATAAAACCTTCTCCAGAGGAATAGAGCTGAAGAC 1629
DB 1483 ACATTTTATATAACAGCCATCCAGATAAAACCTTCTCCAGAGGAATAGAGCTGAAGAC 1542
QY 1630 CAGACTTCAACAGTTGTAACTACAGGAATCAGTATATTACTGACATTTGACGCTGAACAG 1689
DB 1543 CAGACTTCAACAGTTGTAACTACAGGAATCAGTATATTACTGACATTTGACGCTGAACAG 1602
QY 1690 CTGCTTATGTTATCAGGAGCTTGTACCTTTCACTGAGCAGATGATGATGATCTGCT 1749
DB 1603 CTGACTTATGTTCTTATCAGATTAAGGAGATTTGGGCTGAGACAAATGGGGTTTTCTAGA 1662
QY 1750 TTCAACCATCATGGGAGAGGACCAACAGAGTTCTCAGTGTGTTAGGACAGCTCAGCAAGTG 1809
DB 1663 TATACATCATG-----TCACTCTGCAAGCAGGACAAATTTGACTTCCCAAGG 1710
QY 1810 CCAAGCTCCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1869
DB 1711 CTTTGTGAGCCCAAAATTTTCAAGGTTACACATGTTTACCATAACAGAAAGTATTTTACAC 1770
QY 1870 TGGGATCTCCAGATATCCCAATGGAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1929
DB 1771 TGGGATCTCCAGATCTCTGTAATTTTTCATCAATTAACCTTATCACTAATTTTGAATGTGA 1830
QY 1930 TTGGATATAAAAACAGAGCATTTCCAGATAAATAAATAAATAAATAAATAAATAAATAA 1989
DB 1831 AACCAATCCAGAGTATTAATTTTGAAGACATTAACAGAGTTTGTCTCTGCTTATAGG 1890
QY 1990 TTAAGAAATATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2049

Db	1891		TTAAAGAAATACACAAAATACAAAATGAGAGTGGCAGCCTCAACCCACGCTTGAGAGAACT	1950	Db	2965	ACTTCAGGTACTTTTATGTCAGAAATTTTACATCCATGAAGTAACCAATGACTTTGCAAT	3024
Qy	2050	TCTTTGCTGGAAGAAATGACATCTTTTGAGAGAACTTTCAGAGATGAACCGGAATCATCA	2109	Qy	3130	ATGACTGTATCCACAATTTATAGATAAACTGACAATTTCCAGCTACTATATACATTTTCGTTA	3189	
Db	1951	TCTTTGCTGGAAGAAATGACATCTTTTGAGAGAACTTTCAGAGATGAACCGGAATCATCA	2010	Db	3025	ATGACTGTATCCACAATTTATAGATAAACTGACAATTTCCAGCTACTATATACATTTTCGTTA	3084	
Qy	2110	CCTCAAGATGTCGAAGTAAATGATGTTACCGCAGATGAATAAGGTTGAAGTGGTCACCA	2169	Qy	3190	ACAGCAAGTACTCTCAGTTTGGAAATGGGAATAAAAGCAGTGACATCAATGAAGTATACACA	3249	
Db	2011	CCTCAAGATGTCGAAGTAAATGATGTTACCGCAGATGAATAAGGTTGAAGTGGTCACCA	2070	Db	3085	ACAGCAAGTACTCTCAGTTTGGAAATGGGAATAAAAGCAGTGACATCAATGAAGTATACACA	3144	
Qy	2170	CCGAAAGCCCAATGGGATCATTTATGCTTATGAAGTGCTATPATAAAAATATAGATACT	2229	Qy	3250	GATCAAGACATACCTGGAAGGTTTGTGGAAACCTTGACTTACGAATCCATTTTCGTCOAAT	3309	
Db	2071	CCGAAAGCCCAATGGGATCATTTATGCTTATGAAGTGCTATPATAAAAATATAGATACT	2130	Db	3145	GATCAAGACATACCTGGAAGGTTTGTGGAAACCTTGACTTACGAATCCATTTTCGTCOAAT	3204	
Qy	2230	TTATATATGAAGAACACATCAACACAGACATTAATTAAGAACCTTAAGACCTCACCC	2289	Qy	3310	GCAATAAATGTAAGCTGGGTCCACCGGCTCAACCAACCGGTCTAGTCTTCTACATGTT	3369	
Db	2131	TTATATATGAAGAACACATCAACACAGACATTAATTAAGAACCTTAAGACCTCACCC	2190	Db	3205	GCAATAAATGTAAGCTGGGTCCACCGGCTCAACCAACCGGTCTAGTCTTCTACATGTT	3264	
Qy	2290	CTCTATAACATTTCTGTAAGGCTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT	2349	Qy	3370	TCACCTGATCTTACAGCAGACTCTCGCCATGTGAGACCACTCTTGTTTACATATGAGAGA	3429	
Db	2191	CTCTATAACATTTCTGTAAGGCTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT	2250	Db	3265	TCACCTGATCTTACAGCAGACTCTCGCCATGTGAGACCACTCTTGTTTACATATGAGAGA	3324	
Qy	2350	TTACTCTCTGTAAGGACTTCGGAGCTGTGCTGTAGTGACACAGAAAATATCACTTAC	2409	Qy	3430	AGCATATATTTTGTAAATCTGGAATAATACATGATATATATATAAATAATCTCCATCA	3489	
Db	2251	TTACTCTCTGTAAGGACTTCGGAGCTGTGCTGTAGTGACACAGAAAATATCACTTAC	2310	Db	3325	AGCATATATTTTGTAAATCTGGAATAATACATGATATATATATAAATAATCTCCATCA	3384	
Qy	2410	AAAAATATTTCTTTCGGAGATTTGAGCTATCATCTTCCCTCCCAAGTAGTCCCAATGGA	2469	Qy	3490	ACAGAAAGGGATCTCTGATACCTATCTGCTCCAGCTATACATCAAGACTCAAGAAAGAT	3549	
Db	2311	AAAAATATTTCTTTCGGAGATTTGAGCTATCATCTTCCCTCCCAAGTAGTCCCAATGGA	2370	Db	3385	ACAGAAAGGGATCTCTGATACCTATCTGCTCCAGCTATACATCAAGACTCAAGAAAGAT	3444	
Qy	2470	ATCATAAAAATATACAAATTTATCTCAAGAGAAATGAATGGAATCAGGAAAGAACTATA	2529	Qy	3550	GTCCCAAGAACTTCCACCAATAATCAACACTTTTAAAAAACCTTTCTCTACCTCAGTTCTC	3609	
Db	2371	ATCATAAAAATATACAAATTTATCTCAAGAGAAATGAATGGAATCAGGAAAGAACTATA	2430	Db	3445	ATCCCAAGAACTTCCACCAATAATCAACACTTTTAAAAAACCTTTCTCTACCTCAGTTCTC	3504	
Qy	2530	AATPACAACTCTTTAAACCCAAACATTAAGTCTGAAGAAATATACCCAAATATCATTT	2589	Qy	3610	TTATCATGGGATCCCCAGTAAGCCAAATAGTGTGCAATTAATAGTTATGATTTAACTTTA	3669	
Db	2431	AATPACAACTCTTTAAACCCAAACATTAAGTCTGAAGAAATATACCCAAATATCATTT	2490	Db	3505	TTATCATGGGATCCCCAGTAAGCCAAATAGTGTGCAATTAATAGTTATGATTTAACTTTA	3564	
Qy	2590	GAGGTGTCTGTAGTACATGAAGGTGAAGGATTCGGAGTGCTCCCAATAGTACTG	2649	Qy	3670	CAAGGACCAAAATGAAATTTATTTCTTCTTCAATCTCTGATAATTTACATAATTTTGAAGAG	3729	
Db	2491	GAGGTGTCTGTAGTACATGAAGGTGAAGGATTCGGAGTGCTCCCAATAGTACTG	2550	Db	3565	CAAGGACCAAAATGAAATTTATTTCTTCTTCAATCTCTGATAATTTACATAATTTTGAAGAG	3624	
Qy	2650	ACGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAATAACAGTTGTCTGTT	2709	Qy	3730	CTTTTCAACATTTTACATATATATAGCTTTTGTGCGCAGAACTAGAAAAAGGACTTGT	3789	
Db	2551	ACGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAATAACAGTTGTCTGTT	2610	Db	3625	CTTTTCAACATTTTACATATATATAGCTTTTGTGCGCAGAACTAGAAAAAGGACTTGT	3684	
Qy	2710	GTCAAGGTGAAGTTGTATGGCAACCAACCCCTGGAGCCAAATGGAATATCTTTTATTAC	2769	Qy	3790	CCTTCCAGTATTTCTTTTCTTTTACACAGATGATGATGCGCGGTAGGACCTCGCAAAAT	3849	
Db	2611	GTCAAGGTGAAGTTGTATGGCAACCAACCCCTGGAGCCAAATGGAATATCTTTTATTAC	2670	Db	3685	CCTTCCAGTATTTCTTTTCTTTTACACAGATGATGATGCGCGGTAGGACCTCGCAAAAT	3744	
Qy	2770	ACAGTTTATGTCTGGT---AGATCATCATTAATAAACTATTAATGTCTCACTGAAACATCATTTG	2829	Qy	3850	TTGACTTTTAACTCAACTGTACTTTCAGACTTTTGTATGCTGAAATGGAGCCCAAGTCTCTTT	3909	
Db	2671	ACAGTTTATGTCTGGT---AGATCATCATTAATAAACTATTAATGTCTCACTGAAACATCATTTG	2727	Db	3745	TTGACTTTTAACTCAACTGTACTTTCAGACTTTTGTATGCTGAAATGGAGCCCAAGTCTCTTT	3804	
Qy	2830	GAGTTATCAGATTGGAATATATATGTTGAATACAGTGCTTATGTAACAGCTAGCACCCAGA	2889	Qy	3910	CCAGGTGGTATTTGTTTAAAGTATATAGTTTAAAAATTCATGAACTGAAACTGACACTATA	3969	
Db	2728	GAGTTATCAGATTGGAATATATATGTTGAATACAGTGCTTATGTAACAGCTAGCACCCAGA	2787	Db	3805	CCAGGTGGTATTTGTTTAAAGTATATAGTTTAAAAATTCATGAACTGAAACTGACACTATA	3864	
Qy	2890	TTTGGTGATGGGAAAAACAGGAACAAATATCATTTAGCTTTCAACACAGAGGGAGCACCA	2949	Qy	3970	TTATATAAGAAATATATCAGGATTTTAAACTGAAAGCCAAACTTTGTTGGACTGGAAACAGTC	4029	
Db	2788	TTTGGTGATGGGAAAAACAGGAACAAATATCATTTAGCTTTCAACACAGAGGG---ACCA	2844	Db	3865	TATATATAAGAAATATATCAGGATTTTAAACTGAAAGCCAAACTTTGTTGGACTGGAAACAGTC	3924	
Qy	2950	AGGGATCTCCCAAGATGTTTATATGCAAACTCAGTTCTTCATCAATAATCTTTTTC	3009	Qy	4030	AGCACTACTCTATCCGTGTATCTGGTTTCCAAAGTTGGAAATGGCAATCAATTTAGT	4089	
Db	2845	AGGGATCTCCCAAGATGTTTATATGCAAACTCAGTTCTTCATCAATAATCTTTTTC	2904	Db	3925	AGCACTACTCTATCCGTGTATCTGGTTTCCAAAGTTGGAAATGGCAATCAATTTAGT	3984	
Qy	3010	TGGACACCTCTTCAAAACCTTAATGGATTATACAAATTTACTCTGTTTATTTACAGAAAT	3069	Qy	4090	AATGTAGTAAAAATTCACAAACCCCAAGAAATCAGTTCCAGATGTGCTGCAGAAATATGCAAGTGC	4149	
Db	2905	TGGACACCTCTTCAAAACCTTAATGGATTATACAAATTTACTCTGTTTATTTACAGAAAT	2964	Db	3985	AATGTAGTAAAAATTCACAAACCCCAAGAAATCAGTTCCAGATGTGCTGCAGAAATATGCAAGTGC	4044	
Qy	3070	ACTTCAGGTACTTTTATGCGAAATTTTACATCTCATGAACTTAACTCAATGACTTTGCAAT	3129	Qy	4150	ATGGCAACTAGCTGGCAGTCTGTTTATGTAATGGGATTCACCCCAAAAGGCAATGGA	4209	
Db				Db	4045	ATGGCAACTAGCTGGCAGTCTGTTTATGTAATGGGATTCACCCCAAAAGGCAATGGA	4104	

QY 4210 ATAAATACGAGTATATATGTAACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGAT 4269
DB 4105 ATAAATACGAGTATATATGTAACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGAT 4164
QY 4270 CACATGTACACTTTCTATAAGCTTTCTGCCAATACCTCATATGCTCTTTAAAGTAAGGCT 4329
DB 4165 CACATGTACACTTTCTATAAGCTTTCTGCCAATACCTCATATGCTCTTTAAAGTAAGGCT 4224
QY 4330 TCAACCTCAGCTGGTGAAGGTGATGAAGCACATGCCATGTCAGCACACTACCTGAAACA 4389
DB 4225 TCAACCTCAGCTGGTGAAGGTGATGAAGCACATGCCATGTCAGCACACTACCTGAAACA 4284
QY 4390 GTTCCAGGTGTTCCCAAAATATGCTTTTCTGATGTTTCAAGTCAACTAGTCAAAATG 4449
DB 4285 GTTCCAGGTGTTCCCAAAATATGCTTTTCTGATGTTTCAAGTCAACTAGTCAAAATG 4344
QY 4450 ACATGATAGACTGACACTATCTTGGCTACTTTTCAAAATTTACAAAATTTACCACTCA 4509
DB 4345 ACATGATAGACTGACACTATCTTGGCTACTTTTCAAAATTTACAAAATTTACCACTCA 4404
QY 4510 CTTCTGCTCAAAAATGCAAGGATGGAATCCGAAGATGCTGTAATATCAAAAATTT 4569
DB 4405 CTTCTGCTCAAAAATGCAAGGATGGAATCCGAAGATGCTGTAATATCAAAAATTT 4464
QY 4570 CAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGATTTAAAGAAATTTAGA 4629
DB 4465 CAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGATTTAAAGAAATTTAGA 4524
QY 4630 TGGTATAGATTCGAAGTGGTGCAGCACCAATGCTGGCTATGGCAATGCTTCAAACTGG 4689
DB 4525 TGGTATAGATTCGAAGTGGTGCAGCACCAATGCTGGCTATGGCAATGCTTCAAACTGG 4584
QY 4690 ATTTCTACAAAATCTGCTGCTGCCCTCCAGATGGTCCCTCGAAAATGTTTCATGTAGTA 4749
DB 4585 ATTTCTACAAAATCTGCTGCTGCCCTCCAGATGGTCCCTCGAAAATGTTTCATGTAGTA 4644
QY 4750 GCAACATCACTTTTATGATCAGCATAAAGCTGGAGTGAACCTGCTGCTATTAAGTGAACCA 4809
DB 4645 GCAACATCACTTTTATGATCAGCATAAAGCTGGAGTGAACCTGCTGCTATTAAGTGAACCA 4704
QY 4810 ACATGTTATCTGATGATGTCATCAATCGGTAGATAATGATGAATTTAATATATCTTCACTC 4869
DB 4705 ACATGTTATCTGATGATGTCATCAATCGGTAGATAATGATGAATTTAATATATCTTCACTC 4764
QY 4870 AAGTCAAAATGAAGAAATTAACCATAGAAATTAAGATTTAGAAATTTTCAAGGTAT 4929
DB 4765 AAGTCAAAATGAAGAAATTAACCATAGAAATTTAAGATTTAGAAATTTTCAAGGTAT 4824
QY 4930 TCTGTAGTATCATCTGCAATTTACTGGGAAATTTAGTGTGTCATATGTAGAAGGGAAGTCA 4989
DB 4825 TCTGTAGTATCATCTGCAATTTACTGGGAAATTTAGTGTGTCATATGTAGAAGGGAAGTCA 4884
QY 4990 AGTGTGGAATGATGTTACTACTTTAGATACAGCCCAAGAGACCCCTAACACATG 5049
DB 4885 AGTGTGGAATGATGTTACTACTTTAGATACAGCCCAAGAGACCCCTAACACATG 4944
QY 5050 ACATTTTCAAGAGATACAGATGAAGTTTACAAAATTTCAATTAAGTCTTCTCTCTCTCT 5109
DB 4945 ACATTTTCAAGAGATACAGATGAAGTTTACAAAATTTCAATTAAGTCTTCTCTCTCTCTCT 5004
QY 5110 CAACTTAATGGAATATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTACT 5169
DB 5005 CAACTTAATGGAATATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTACT 5064
QY 5170 GCTGTCCAGATTCACAACTCAGTATATACAGAAACCAACATTTGTCATGTCATG 5229
DB 5065 GCTGTCCAGATTCACAACTCAGTATATACAGAAACCAACATTTGTCATGTCATG 5124
QY 5230 CTAGAAGGCTAAAGGTGACATACATCAATATCAGTGTGTTTACCGAGTCAATAGTGT 5289
DB 5125 CTAGAAGGCTAAAGGTGACATACATCAATATCAGTGTGTTTACCGAGTCAATAGTGT 5184

QY 5290 GGTGAGGTCCAAAGGTTCCGATGAGAAATAACCATGGATATCAAGCTCCAGCACGACCA 5349
DB 5185 GGTGAGGTCCAAAGGTTCCGATGAGAAATAACCATGGATATCAAGCTCCAGCACGACCA 5244
QY 5350 AAAACCAAAACCAACCCCTATTTATGATGCCACAGGAAATGCTTCTGACTTCAACAA 5409
DB 5245 AAAACCAAAACCAACCCCTATTTATGATGCCACAGGAAATGCTTCTGACTTCAACAA 5304
QY 5410 ATTACAATCAGAAATGCCAATATGTTTACTACAGTGATGATCATGGACCAATAAAAAATGTA 5469
DB 5305 ATTACAATCAGAAATGCCAATATGTTTACTACAGTGATGATCATGGACCAATAAAAAATGTA 5364
QY 5470 CAAAGTCTTGGACAGAAACAGAGCTCAGCATGATGGAATGTAACAAAGTGGTATGAT 5529
DB 5365 CAAAGTCTTGGACAGAAACAGAGCTCAGCATGATGGAATGTAACAAAGTGGTATGAT 5424
QY 5530 GCATATTTTAAATAAGACAGGCCATATTTTACAATGAAGGCTTTTCTTAACCTCCATGT 5589
DB 5425 GCATATTTTAAATAAGACAGGCCATATTTTACAATGAAGGCTTTTCTTAACCTCCATGT 5484
QY 5590 ACAGAAGGAAAGACAAAGTTTATGTCGCAATGAAGAAATCTACATCATAGGTGCTGATAAT 5649
DB 5485 ACAGAAGGAAAGACAAAGTTTATGTCGCAATGAAGAAATCTACATCATAGGTGCTGATAAT 5544
QY 5650 GCATGATGATTTCTGGCAATGAAGACAAATTTGCAATGGACCACTGAAACCAAAAAG 5709
DB 5545 GCATGATGATTTCTGGCAATGAAGACAAATTTGCAATGGACCACTGAAACCAAAAAG 5604
QY 5710 CAATCTTATTTAAATTTAGAGCTACAAATATTTATGGGACAAATTTACTGACTCTGATAT 5769
DB 5605 CAATCTTATTTAAATTTAGAGCTACAAATATTTATGGGACAAATTTACTGACTCTGATAT 5664
QY 5770 TCTGACCTCTTTAAGAATTTAGGGGAAGGACTTTTCAAGAAAGAACCGTAGAGATCATTTCTT 5829
DB 5665 TCTGACCTCTTTAAGAATTTAGGGGAAGGACTTTTCAAGAAAGAACCGTAGAGATCATTTCTT 5724
QY 5830 TCGGTCACTTTGTGATTCCTTCAATATTTCTCTTGGAAACAGCTATTTTGTGATTTGCA 5889
DB 5725 TCGGTCACTTTGTGATTCCTTCAATATTTCTCTTGGAAACAGCTATTTTGTGATTTGCA 5784
QY 5890 AGAATTCGACAGACAGAGAAAGAGTGGCACATCTCTCCTCAGAGATCAGAAATTTAT 5949
DB 5785 AGAATTCGACAGAGAGAGAGTGGCACATCTCTCCTCAGAGATCAGAAATTTAT 5844
QY 5950 GACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTGAAGGACGAG 6009
DB 5845 GACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTGAAGGACGAG 5904
QY 6010 AGATTAACCGCGCCCAATAGCAAGAAATCTCTTCTGCAACATGTTGAAGAGCTTTGCACA 6069
DB 5905 AGATTAACCGCGCCCAATAGCAAGAAATCTCTTCTGCAACATGTTGAAGAGCTTTGCACA 5964
QY 6070 AACACACCTTAAAGTTTCAAGAGAAATTTTCCGAAATTTACCAAAATTTCTTCAGGATCTT 6129
DB 5965 AACACACCTTAAAGTTTCAAGAGAAATTTTCCGAAATTTACCAAAATTTCTTCAGGATCTT 6024
QY 6130 TCTTCAACTGATGCTGATCTGCTTGGAAATAGCAAAAAACCGTTTCCCAACATAAAA 6189
DB 6025 TCTTCAACTGATGCTGATCTGCTTGGAAATAGCAAAAAACCGTTTCCCAACATAAAA 6084
QY 6190 CCAATATAATAATAACAGATGAAAGCTGATGATGACGCTAGTGTTCAGGTTCCGAT 6249
DB 6085 CCAATATAATAATAACAGATGAAAGCTGATGATGACGCTAGTGTTCAGGTTCCGAT 6141
QY 6250 TATATTTATGCCAGCTTATATTTCTGGTATTTATGTCGCAATTTATTTCTACTCTCAA 6309
DB 6142 TATATTTATGCCAGCTTATATTTCTGGTATTTATGTCGCAATTTATTTATTTCTACTCAA 6201
QY 6310 GGTCCACTACCAAGAAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACCCAGGGCAAAA 6369
DB 6202 GGTCCACTACCAAGAAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACCCAGGGCAAAA 6261
QY 6370 ACATTTAGTAATGCTAACACAGTGTGTTTGAAGAAAGGACGATCAGATGCCATCAGTATGG 6429

Db 1663 TATACATCATG-----TCATCTGCAAGCAGGACAAATTTTGACTTCCCCAGGC 1710
QY 1810 CCAAGCTCCATTAAATATATAAATAATATATAGTCTCTTCACTATTTTGTATAT 1869
Db 1711 CTTTGTGAGCCCAAAATTTGAGAGTTACATGTTTACCATTAACAGAAATTTTTTACAC 1770
QY 1870 TGGGATCCTCCAGAAATATCCCAATGGAATAATTAACCTCATATACGAATTTATCGAATGGA 1929
Db 1771 TGGGATCCTCCAGATCCTGTATTTTTCATCATTTACCTTATCATATTTTGGATGTTGAA 1830
QY 1930 TTGGATACAAACAGAGCAATTCAGATAAATACCATAGATAAACAGCTTCTCATACAGGG 1989
Db 1831 AACCANTCCAAGNGATTAATTTTAAGGACATTAACAGTTTGTCTCTTGTCTTATAGG 1890
QY 1990 TTAAGAAATACACAAAATACAAAATGAGAGTGCGCAGCTCAACCCACGATGAGAAATG 2049
Db 1891 TTAAGAAATACACAAAATACAAAATGAGAGTGCGCAGCTCAACCCACGTTGAGAAATG 1950
QY 2050 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTTCAGAGATGAACCGGAATCATCA 2109
Db 1951 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTTCAGAGATGAACCGGAATCATCA 2010
QY 2110 CTTCAAGATGTCGAAGTAAATGTATGTATCCGAGATGAATAAGGTTGAAGTGGTCACCA 2169
Db 2011 CTTCAAGATGTCGAAGTAAATGTATGTATCCGAGATGAATAAGGTTGAAGTGGTCACCA 2070
QY 2170 CCGGAAAGCCCAATCGGATCAATTTATGTCTTATGAAGTGTCTATATAAAATATAGATCT 2229
Db 2071 CCGGAAAGCCCAATCGGATCAATTTATGTCTTATGAAGTGTCTATATAAAATATAGATCT 2130
QY 2230 TTATATATGAAGAACATCAACAGACAGACATTAATTAAGNACTTTRAGACTCACCC 2289
Db 2131 TTATATATGAAGAACATCAACAGACAGACATTAATTAAGNACTTTRAGACTCACCC 2190
QY 2290 CTCTATAACATTTCTGTAGGCTCTACACAGATTTGGTTCATGGCAATCAGGTATCTTCT 2349
Db 2191 CTCTATAACATTTCTGTAGGCTCTACACAGATTTGGTTCATGGCAATCAGGTATCTTCT 2250
QY 2350 TTACTCTCTGTAAAGGACTCGGAGACTGTGCGCTGATGTGACAGAGAAATATCACTTAC 2409
Db 2251 TTACTCTCTGTAAAGGACTCGGAGACTGTGCGCTGATGTGACAGAGAAATATCACTTAC 2310
QY 2410 AAAAATATTTCTCTCGAGAGATTGAGCTATCATCTCTCCCGCAGTAGTCCCAATGGA 2469
Db 2311 AAAAATATTTCTCTCGAGAGATTGAGCTATCATCTCTCCCGCAGTAGTCCCAATGGA 2370
QY 2470 ATCATAAAAAATATACAAATTTATCTCAAGAGAAAGTAATGGAAATGAGGAAAGAACTATA 2529
Db 2371 ATCATACAAAATATACAAATTTATCTCAAGAGAAAGTAATGGAAATGAGGAAAGAACTATA 2430
QY 2530 AATPACAACTCTTTAAACCCAAAACATTAAGTACTCAAGAAATATATCCCAATATATCAT 2589
Db 2431 AATPACAACTCTTTAAACCCAAAACATTAAGGCTCAAGAAATATATCCCAATATATCAT 2490
QY 2590 GAGGTCTGTCTAGTACACTGAAGGTGAAGGAGTTGCGAGTGCTCCCATAGTACTG 2649
Db 2491 GAGGTCTGTCTAGTACACTGAAGGTGAAGGAGTTGCGAGTGCTCCCATAGTACTG 2550
QY 2650 ACGGAGGAGATGCTCTGTATCTCCCTCAAGACTTCTCTGTAAGAAAGTGTCTGGT 2709
Db 2551 ACGGAGGAGATGCTCTGTATCTCCCTCAAGACTTCTCTGTAAACAGTTGTCTGGT 2610
QY 2710 GTCAAGGTGAAGTTGTATGGCAACCAACCCCTGGAGCCAAATGAATATATCTTTTATAC 2769
Db 2611 GTCAAGGTGAAGTTGTATGGCAACCAACCCCTGGAGCCAAATGAATATATCTTTTATAC 2670
QY 2770 ACAGTTTATGTCGGAATAGATCATCATTAATAAATATTAATGTCATGGAACATCATTTG 2829
Db 2671 ACAGTTTATGTCGGAATAGATCATCATTAATAAATATTAATGTCATGGAACATCATTTG 2727
QY 2830 GAGTTATCAGATTTGGATTAATATGTTGAATACAGTGCTTATGTAAACAGCTAGCACCAGA 2889

Db 2728 GAGTTATCAGATTTGGATTAATAATGTTGAATAACAGTGTCTTATGTAAACAGCTAGCACCAGA 2787
QY 2890 TTTTGTGTGATGGAAACAGGAGCAATATCATTTAGCTTTCAAAACACCCAGAGAGGAGCACCA 2949
Db 2788 TTTTGTGTGATGGAAACAGGAGCAATATCATTTAGCTTTCAAAACACCCAGAGAGG---ACCA 2844
QY 2950 AGCGATCCTCCCAAAAGATGTTTATGTATGCAAACTCTCAGTTCTTCATCAATAATTTCTTTTC 3009
Db 2845 AGCGATCCTCCCAAAAGATGTTTATGTATGCAAACTCTCAGTTCTTCATCAATAATTTCTTTTC 2904
QY 3010 TGGACACTCTCTTCAAAACCTTAATGGGATATACAAATATCTCTGTGTTTATACAGAAAT 3069
Db 2905 TGGACACTCTCTTCAAAACCTTAATGGGATATACAAATATCTCTGTGTTTATACAGAAAT 2964
QY 3070 ACTTCAGGTACTTTTATGCGAATTTTACACTCCATGAACCTTAACCAATGACCTTTGACAAT 3129
Db 2965 ACTTCAGGTACTTTTATGCGAATTTTACACTCCATGAAGTAACCAATGACCTTTGACAAT 3024
QY 3130 ATGACTGTATCCCAATATATAGATAAACTGACAAATATTCAGCTACTATACATTTTGGTTA 3189
Db 3025 ATGACTGTATCCCAATATATAGATAAACTGACAAATATTCAGCTACTATACATTTTGGTTA 3084
QY 3190 ACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGTGAATCATTTGAAGTATACACA 3249
Db 3085 ACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGTGAATCATTTGAAGTATACACA 3144
QY 3250 GATCAGACATACCTGAAAGGTTCTTGGAAACCTGACTTACGAATCCATTTTCGCAACT 3309
Db 3145 GATCAGACATACCTGAAAGGTTCTTGGAAACCTGACTTACGAATCCATTTTCGCAACT 3204
QY 3310 GCAATAAATGTAAGCTGGGTCCCAACGGCTCAACCAAAACGGTCTAGTCTTTCTACTATGTT 3369
Db 3205 GCAATAAATGTAAGCTGGGTCCCAACGGCTCAACCAAAACGGTCTAGTCTTTCTACTATGTT 3264
QY 3370 TCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACTCTTGTGTTACATATGAGAGA 3429
Db 3265 TCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACTCTTGTGTTACATATGAGAGA 3324
QY 3430 AGCATATATTTGATAAATCTGGAAATAACACTGATATATATAATAATAATAATAATAATAATA 3489
Db 3325 AGCATATATTTGATAAATCTGGAAATAACACTGATATATATAATAATAATAATAATAATAATA 3384
QY 3490 ACAGAAAAGGGATTTCTGTATACCTATATCGCCAGCTATATACATCAAGACTGAAGAAGAT 3549
Db 3385 ACAGAAAAGGGATTTCTGTATACCTATATCGCCAGCTATATACATCAAGACTGAAGAAGAT 3444
QY 3550 GTCCCAAGAACTTCACCAATAATCAACACTTTTAAACCTTTCTCTACCTCAGTTCTC 3609
Db 3445 ATCCCAAGAACTTCACCAATAATCAACACTTTTAAACCTTTCTCTACCTCAGTTCTC 3504
QY 3610 TTATCATGCGATCCCCAGTAAAGCCAAATGTTGCAATAATAAGTTATGATTTAACTTTA 3669
Db 3505 TTATCATGCGATCCCCAGTAAAGCCAAATGTTGCAATAATAAGTTATGATTTAACTTTA 3564
QY 3670 CAAGGACCAAAATGAAAATTTCTTTTCACTTCTGTATAATTTACATAATATTTGGAAGAG 3729
Db 3565 CAAGGACCAAAATGAAAATTTCTTTTCACTTCTGTATAATTTACATAATATTTGGAAGAG 3624
QY 3730 CTTTCAACATTTACATATATAGCTTTTGTGCGCGCAAGAACTAGAAAAGGACTTGGT 3789
Db 3625 CTTTCAACATTTACATATATAGCTTTTGTGCGCGCAAGAACTAGAAAAGGACTTGGT 3684
QY 3790 CTTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCCGCTAGCTCCACCAAAAT 3849
Db 3685 CTTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCCGCTAGCTCCACCAAAAT 3744
QY 3850 TTGACTTTAATCAACTGTACTTTCAGACTTTTGTATGGCTGAAAATGGAGCCCAAGTCTCTCT 3909
Db 3745 TTGACTTTAATCAACTGTACTTTCAGACTTTTGTATGGCTGAAAATGGAGCCCAAGTCTCTCT 3804
QY 3910 CCAGGTGGTATTTGTTAAAGTATATAGTTTAAATTTCAATGAACTGAAACTGACACTATA 3969
Db 3805 CCAGGTGGTATTTGTTAAAGTATATAGTTTAAATTTCAATGAACTGAAACTGACACTATA 3864

QY 3970 TATTATAAGAAATATATACAGATTTAAAACTGAAGCCAAACTTGTGTGGACTGGAAACAGTGC 4029
DB 3865 TATTATAAGAAATATATACAGATTTAAAACTGAAGCCAAACTTGTGTGGACTGGAAACAGTGC 3924
QY 4030 AGCACTTACTCTATCCGTGTATCTCGTTCACCAAAAGTTGGAATGGAATGCAATCAATTTAGT 4089
DB 3925 AGCACTTACTCTATCCGTGTATCTCGTTCACCAAAAGTTGGAATGGAATGCAATCAATTTAGT 3984
QY 4090 AATGTAGTAAATTTCAAAACCCCAAGAAATCAGTTTCCAGATGTGTGTGAGAAATATGCAAGTGC 4149
DB 3985 AATGTAGTAAATTTCAAAACCCCAAGAAATCAGTTTCCAGATGTGTGTGAGAAATATGCAAGTGC 4044
QY 4150 ATGGCAACTAGCTGGCAGTCAAGTTTATGTAAGAAATGGAATGCCAACCCAAAGGCAAAATGGA 4209
DB 4045 ATGGCAACTAGCTGGCAGTCAAGTTTATGTAAGAAATGGAATGCCAACCCAAAGGCAAAATGGA 4104
QY 4210 ATAAATTAACCCAGTATATGTGAACAGTTGAAAGGAAATTCACAAAGGTTTCTCCCAAGAT 4269
DB 4105 ATAAATTAACCCAGTATATGTGAACAGTTGAAAGGAAATTCACAAAGGTTTCTCCCAAGAT 4164
QY 4270 CACATGTACACTTTTCAAAAGCTTCTTGCCAAATACCTCATATGTCTTTAAAGTAAGAGCT 4329
DB 4165 CACATGTACACTTTTCAAAAGCTTCTTGCCAAATACCTCATATGTCTTTAAAGTAAGAGCT 4224
QY 4330 TCAACTCAGCTGGTGAAGGTGAAGAAAGCAATGCCATGTGAGCACTACTCTGAAACA 4389
DB 4225 TCAACTCAGCTGGTGAAGGTGAAGAAAGCAATGCCATGTGAGCACTACTCTGAAACA 4284
QY 4390 GTTCCAGTGTCCCAAAATATGCTTTTCTGTGATGTTCAGTCAACTAGTGCACATTTG 4449
DB 4285 GTTCCAGTGTCCCAAAATATGCTTTTCTGTGATGTTCAGTCAACTAGTGCACATTTG 4344
QY 4450 ACATGTGATGAAGCTGACACTATCTTGTGCTACTTTCAAAATTAACAAATTAACCACTCAA 4509
DB 4345 ACATGTGATGAAGCTGACACTATCTTGTGCTACTTTCAAAATTAACAAATTAACCACTCAA 4404
QY 4510 CTTGTGCTCAAAAATGCAAAAGATGGGAATCCGAAGAAATGTTTGAATPATCAAAAATTT 4569
DB 4405 CTTGTGCTCAAAAATGCAAAAGATGGGAATCCGAAGAAATGTTTGAATPATCAAAAATTT 4464
QY 4570 CAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGATTTAAAGAAATTTAGA 4629
DB 4465 CAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGATTTAAAGAAATTTAGA 4524
QY 4630 TGGTATAGATTCGAAGTGGCTGCAGCACTATGCTGGCTATGGCAATGCTTCAAACTGG 4689
DB 4525 TGGTATAGATTCGAAGTGGCTGCAGCACTATGCTGGCTATGGCAATGCTTCAAACTGG 4584
QY 4690 ATTTCTCAAAAATCTGCTGGCTCCAGATGGTCTCTGAAATGTTTCAATGTAGTA 4749
DB 4585 ATTTCTCAAAAATCTGCTGGCTCCAGATGGTCTCTGAAATGTTTCAATGTAGTA 4644
QY 4750 GCAACATCACTTTTAGCATCAGCATTAAGCTGGAGTGAACCTGCTGTCTATTAAGTGA 4809
DB 4645 GCAACATCACTTTTAGCATCAGCATTAAGCTGGAGTGAACCTGCTGTCTATTAAGTGA 4704
QY 4810 ACATGTTATCTGATGTCAAAATCGGTAGATTAATGATGAATTTAAATATATCTTCTATC 4869
DB 4705 ACATGTTATCTGATGTCAAAATCGGTAGATTAATGATGAATTTAAATATATCTTCTATC 4764
QY 4870 AAGTCAAAATGAAGAAATTAACCAATAGAAATTTAGAAATTTTCAAGGATAT 4929
DB 4765 AAGTCAAAATGAAGAAATTAACCAATAGAAATTTAGAAATTTTCAAGGATAT 4824
QY 4930 TCTGTAGTCACTAGCTTTTACTGGGAACATTTAGTGTGCAATATGTAAGGGAAGTCA 4989
DB 4825 TCTGTAGTCACTAGCTTTTACTGGGAACATTTAGTGTGCAATATGTAAGGGAAGTCA 4884
QY 4990 AGTGCTGAAATGATTTGTTACTACTTTTGAATCAGCCCCCAAGGACCCCACTAACACATG 5049
DB 4885 AGTGCTGAAATGATTTGTTACTACTTTTGAATCAGCCCCCAAGGACCCCACTAACACATG 4944

QY 5050 ACATTTTCAGAAATACCAAGTATCAAAATTTCAATTAACGTTCTTCTCTCTTCT 5109
DB 4945 ACATTTTCAGAAATACCAAGTATCAAAATTTCAATTAACGTTCTTCTCTCTTCT 5004
QY 5110 CAACCTTAATGGAATATCAAGTATATCAAGTCTCTGTTTACCGAGAAGATGATCTTACT 5169
DB 5005 CAACCTTAATGGAATATCAAGTATATCAAGTCTCTGTTTACCGAGAAGATGATCTTACT 5064
QY 5170 GCTGCTCAGATTCACAACTCAGTATATATACAGAAACCAACACATTCGTCATTTGCAATG 5229
DB 5065 GCTGCTCAGATTCACAACTCAGTATATATACAGAAACCAACACATTCGTCATTTGCAATG 5124
QY 5230 CTAGAAGGACTAAAAGGTGGACATATACATCAATATCAGTGTGTTTACGAGTCAATAGTCT 5289
DB 5125 CTAGAAGGACTAAAAGGTGGACATATATCAATATCAGTGTGTTTACGAGTCAATAGTCT 5184
QY 5290 GGTGCAAGTCCAAAGTTCAGATGAGAAATTAACATGATATCAAAAGTCCAGCAACCA 5349
DB 5185 GGTGCAAGTCCAAAGTTCAGATGAGAAATTAACATGATATCAAAAGTCCAGCAACCA 5244
QY 5350 AAAACCAAAACCAACCTTATTTATGATGCCACAGGAAAACTGCTGTGTGATTTCAACAA 5409
DB 5245 AAAACCAAAACCAACCTTATTTATGATGCCACAGGAAAACTGCTGTGTGATTTCAACAA 5304
QY 5410 ATTCAATCAGATGCGCAATATGTTTACTACATGATGATCATGAGCAATATAAAATGTAT 5469
DB 5305 ATTCAATCAGATGCGCAATATGTTTACTACATGATGATCATGAGCAATATAAAATGTAT 5364
QY 5470 CAAGTGTCTGCGACAGAAACAGAGCTCAGCATGATGGAATGTAACAAAGTGTGATGAT 5529
DB 5365 CAAGTGTCTGCGACAGAAACAGAGCTCAGCATGATGGAATGTAACAAAGTGTGATGAT 5424
QY 5530 GCATATTTTAAATGAAGCCATATTTTAAATGAAGGCTTTCTCAACCTCCATGAT 5589
DB 5425 GCATATTTTAAATGAAGCCATATTTTAAATGAAGGCTTTCTCAACCTCCATGAT 5484
QY 5590 ACAGAAAGGAAGACAAAGTTTATGTCGCAATGAAGAAATCTAATCATAGGTGTGATAAT 5649
DB 5485 ACAGAAAGGAAGACAAAGTTTATGTCGCAATGAAGAAATCTAATCATAGGTGTGATAAT 5544
QY 5650 GCATGCATGATTCCTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAACCAAAAG 5709
DB 5545 GCATGCATGATTCCTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAACCAAAAG 5604
QY 5710 CAATACCTTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATTA 5769
DB 5605 CAATACCTTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATTA 5664
QY 5770 TCTGACCTCTGTTAAGACTTTTAGGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCATCTT 5829
DB 5665 TCTGACCTCTGTTAAGACTTTTAGGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCATCTT 5724
QY 5830 TCCGTCATCTTGTGTATCTTTCAATATTTCTTCTGGAACAGCTATTTTTCGATTTGCA 5889
DB 5725 TCCGTCATCTTGTGTATCTTTCAATATTTCTTCTGGAACAGCTATTTTTCGATTTGCA 5784
QY 5890 AGAATTTGCAAGACAGAAAGGCTGGCACTACTCTCTCAGGATGACAGAAATTTAT 5949
DB 5785 AGAATTTGCAAGACAGAAAGGCTGGCACTACTCTCTCAGGATGACAGAAATTTAT 5844
QY 5950 GACACTAAATGGAAGCTGGATCAGCTCATCAGTGGGAGAGCTGGAACTGAAGGACGAG 6009
DB 5845 GACACTAAATGGAAGCTGGATCAGCTCATCAGTGGGAGAGCTGGAACTGAAGGACGAG 5904
QY 6010 AGATTAACGGCGCCCAATGAAGCAAAATCTTCTGCAACATGTTTGAAGAGCTTTGCA 6069
DB 5905 AGATTAACGGCGCCCAATGAAGCAAAATCTTCTGCAACATGTTTGAAGAGCTTTGCA 5964
QY 6070 AACCAACCTTAAAGTTTCAAGAAATTTTTCGGAATTTACCAAAATTTCTTCAGGATCTT 6129
DB 5965 AACCAACCTTAAAGTTTCAAGAAATTTTTCGGAATTTACCAAAATTTCTTCAGGATCTT 6024
QY 6130 TCTTCAACTGATGCTGATCTGCTTTGGAATGAGAGCAAAACCGTTTCCCAACATAAAA 6189

Db 6025 TCTTCACTGATGCTGCTGCTGGAATAGACGAAAAACCGCTTCCCAACATATAA 6084
QY 6190 CCATATAATAATAAACAAGAGTAAGCTGATAGCTGAGCGCTAGTGTTCAGGTTCCGAT 6249
Db 6085 CCAT---ATAATAATAAACAAGAGTAAGCTGATAGCTGAGCGCTAGTGTTCAGGTTCCGAT 6141
QY 6250 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCGAAATGAATTTATTCCTACTCAA 6309
Db 6142 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCGAAATGAATTTATTCCTACTCAA 6201
QY 6310 GGTCCATACCAAGCAAGCTGCGAGATTTTTCGAGAATGGTGTGGGAACCAAGGCGCAAAA 6369
Db 6202 GGTCCATACCAAGCAAGCTGCGAGATTTTTCGAGATGGTGTGGGAACCAAGGCGCAAAA 6261
QY 6370 ACATTAGTAATCTAAACAGTGTGTTTGAAGGAGCGGATCAGATGCCATCAGTATTGG 6429
Db 6262 ACATTAGTAATCTAAACAGTGTGTTTGAAGGAGCGGATCAGATGCCATCAGTATTGG 6321
QY 6430 CCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTCAATTTACAAAGCTAATGGAG 6489
Db 6322 CCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTCAATTTACAAAGCTAATGGAG 6381
QY 6490 GATGTTCAATAGATTGGAATCTAGGATCTGAAATTTGAAGGATGGGATTTGCATG 6549
Db 6382 GATGTTCAATAGATTGGAATCTAGGATCTGAAATTTGAAGGATGGGATTTGCATG 6441
QY 6550 ACTGTTGCAAGTGAATCTTTACTGCTGCGCAGAGCATGCGGTTCTGAGAACACGCGC 6609
Db 6442 ACTGTTGCAAGTGAATCTTTACTGCTGCGCAGAGCATGCGGTTCTGAGAACACGCGC 6501
QY 6610 CCTCTAATTTCACTTTGTGAAGTTGTTGAGCAACGAGGACATGACACCACTATG 6669
Db 6502 CCTCTAATTTCACTTTGTGAAGTTGTTGAGCAACGAGGACATGACACCACTATG 6561
QY 6670 ATTGTTCACTGAGTGTGGAGTTGGAAGAACTGGAGTTTATTTGCTCTGACCAATTTA 6729
Db 6562 ATTGTTCACTGAGTGTGGAGTTGGAAGAACTGGAGTTTATTTGCTCTGACCAATTTA 6621
QY 6730 ACAACAATATAAATGACATGATTTTGTGGATATATATGGAATGAGTGAATGAGTGA 6789
Db 6622 ACAACAATATAAATGACATGATTTTGTGGATATATATGGAATGAGTGAATGAGTGA 6681
QY 6790 AGTGAAGAAATGTGATGTCGAGATCTGCGACAGTATATCTTTTACACCACTGCAAT 6849
Db 6682 AGTGAAGAAATGTGATGTCGAGATCTGCGACAGTATATCTTTTACACCACTGCAAT 6741
QY 6850 CTGGATCTCTTATCAATAAGGAAAGTATCAGCCCATCTGTTTGTGTTAACTATTTCAGCA 6909
Db 6742 CTGGATCTCTTATCAATAAGGAAAGTATCAGCCCATCTGTTTGTGTTAACTATTTCAGCA 6801
QY 6910 CTTGAGAAGATGCACTCTTTGAGCGCATGGAA---GGTGAATGTTGAGCTTGAATGGGAA 6966
Db 6802 CTTGAGAAGATGCACTCTTTGAGCGCATGGAAAGGTGGTGTGATGTTGAGCTTGAATGGGAA 6861
QY 6967 GAAACCACTATGTAATATTTCAGCAACCAAGGATCAATTTGGAAGAGATTTTAAATCCCA 7026
Db 6862 GAAACCACTATGTAATATTTCAGCAACCAAGGATCAATTTGGAAGAGATTTTAAATCCCA 6921
QY 7027 GGGGGCAAAATGTAACCCCTCATTTCTTCCAAATTTGAAATGTGCAACCTTAAAGAAATATCT 7086
Db 6922 GGGGGCAAAATGTAACCCCTCATTTCTTCCAAATTTGAAATGTGCAACCTTAAAGAAATATCT 6981
QY 7087 ATGCTTCTCTCAC 7099
Db 6982 ATGCTTCTCTCAC 6994

RESULT 9
ADH41612
ID ADH41612 standard; DNA; 6881 BP.
XX
AC ADH41612;

XX 25-MAR-2004 (first entry)
DT Novel human nucleic acid NOV15a.
DE ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;
KW anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;
KW neotropic; antiparkinsonian; antiasthmatic; antiinfertility;
KW cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
KW AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX Homo sapiens.
OS
XX WO2003102159-A2.
XX 11-DEC-2003.
XX 04-JUN-2003; 2003WO-US017573.
XX 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386864P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387081P.
PR 07-JUN-2002; 2002US-0387083P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387866P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387686P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.

PR 23-OCT-2002; 2002US-0420627P.
 PR 23-OCT-2002; 2002US-0420718P.
 PR 24-OCT-2002; 2002US-0420852P.
 PR 31-OCT-2002; 2002US-0422750P.
 PR 01-NOV-2002; 2002US-0423095P.
 PR 05-NOV-2002; 2002US-0423748P.
 XX (CURA-) CURAGEN CORP.
 XX
 PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
 PI Burgess CE, Casman SJ, Catterton E, Dhanabai M, Edinger SR;
 PI Ellerman K, Eitenberg S, Gangolli EA, Gerlach VL, Gorman L;
 PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
 PI Khramtsov NV, Larochelele WJ, Li L, Liang H, Low K, Macdougall JR;
 PI MacLachlan T, Maliyankar UM, McQueeney K, Mezick AJ, Miller CE;
 PI Millet I, Padigaru M, Petturajan M, Peyman J, Qian X, Raschelli L;
 PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
 PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CM, Voss EZ;
 PI Wolenc AR, Zhong M, Zhong H;
 XX
 DR WPI; 2004-053467/05.
 DR P-PSDB; ADH41613.
 XX
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
 PT pharmacogenomics.
 XX
 PS Claim 20; SEQ ID NO 165; 1503pp; English.
 XX
 CC The invention relates to 566 new isolated human polypeptides and their
 CC encoding genes, sequences that are at least 95% identical to these or
 CC sequences comprising one or more conservative substitutions in these. The
 CC polypeptide, polynucleotide and antibodies against the polypeptides are
 CC useful in diagnosing, treating or preventing NOVX-associated disorders,
 CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
 CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
 CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
 CC The nucleic acids are further used as hybridization probes, in chromosome
 CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
 CC polypeptides are also useful as vaccines. This sequence represents an
 CC example of the nucleic acid sequence of the invention.
 XX
 SQ Sequence 6881 BP; 2297 A; 1362 C; 1301 G; 1921 T; 0 U; 0 Other;
 Query Match 89.2%; Score 6340; DB 12; Length 6881;
 Best Local Similarity 96.2%; Pred. No. 0;
 Matches 6657; Conservative 0; Mismatches 215; Indels 48; Gaps 14;
 QY 91 AAAAAATGTAATAAGATGGATTTCTTATCATTTTCTTTTATTTGAGCTTCA 150
 DB 1 AAAAAATGTAATAAGATGGATTTCTTATCATTTTCTTTTATTTTATTTGAGCTTCA 60
 QY 151 GAGACACAGGTGATGTTTCCAAATGTCGTTCTCTGTTAGTAGATTAACCAATCTCT 210
 DB 61 GAGACACAGGTAGATGTTTCCAAATGTCGTTCTCTGTTAGTAGATTAACCAATCTCT 120
 QY 211 TCAATTTCTACACATACACCTCCTCCTGTTTCTGTTAGTAGATTAACCAATCTCAACAAA 270
 DB 121 TCAATTTCTACACATACACCTCCTCCTGTTTCTGTTAGTAGATTAACCAATCTCAACAAA 174
 QY 271 CCAGGGCTCTCAGTCTTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTGCTCT 330
 DB 175 CCAGGGCTCTCAGTCTTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTTCTTGCTCT 234
 QY 331 TGAATATACACCTTAATCCAAATGGAAGGATTAATCTTCAATGTTCAAAATATAGGAA 390
 DB 235 TGAATATACACCTTAATCCAAATGGAAGGATTAATCTTCAATGTTCAAAATATAGGAA 294
 QY 391 GTTTCCTCGTGGATGAACAGTATATACACAGTCAGATCAAGCCGACAGCTTGGAA 450
 DB 295 GTTTCCTCGTGGATGAACAGTATATACACAGTCAGATCAAGCCGACAGCTTGGAA 354

QY 451 GTTCTCTTACTTAATCTTAATCTCTCGAACACATATGAAATTAAGTTCCTGCTGTAAC 510
 DB 355 GTTCTCTTACTTAATCTTAATCTCTCGAACACATATGAAATTAAGTTCCTGCTGTAAC 414
 QY 511 AGTGTCTGGCATTGGAGTGTGTTAGTGTATCCATTTCTTCCAAACTCAGAAAAGTGTCTCA 570
 DB 415 AGTGTCTGGCATTGGAGTGTGTTAGTGTATCCATTTCTTCCAAACTCAGAAAAGTGTCTCA 474
 QY 571 GGAAAGTGGTGAATCTCA CAGTTGAGGCTTCAACGGCTTTCAGCAGTTAAGCTGATTTGG 630
 DB 475 GGAAAGTGGTGAATTTTACAGGCTGAGGCTGTC--CCGTTACGAG-TAAGCTGATGTGG 531
 QY 631 TATTTACCTCGCAACCAATGGCAAAATACAGCTTCAAGATAGTGTCAAAATGCTCC 690
 DB 532 TA-TACCTCGCAACCAAAA-AAAAAAATTACAGCTTCAAGATAGTGTCAAGCAATAAC 588
 QY 691 AGAAGTGGGATAGTAGTGAAGATGTCTCAATCAGAGTAGAGGACATTTTGAAGTGGGAAA 750
 DB 589 AGAAGTGGGATAGTAGTGAAGAGTGTCTCATCAGAGTGGAGTGCAATTTTNAAGTGTCTCC 648
 QY 751 TTGCC--AGAAATGCAATGAGAAATAGTGAATCTTTTATGGAGTACAGCCAGCCCTTCT 807
 DB 649 CTTCCTTTGCACTGCAACAGAGAAATAGTGAATCTTTTATGGAGTACAGCCAGCCCTTCT 708
 QY 808 CCAACCTTTGTTAGAGTTACACCTTCATCGCTGACACACATTCATCAAGACAGTTGACA 867
 DB 709 CCAACCTTTGTTAGAGTTACACCTTCATCGCTGACACACATTCATCAAGACAGTTGACA 768
 QY 868 CAGAATGAGTACAGCTCTGTGGAAAGAGCCTATCAGTTTGTAGTGAACACTTGAGA 927
 DB 769 CAGAATGAGTACAGCTCTGT--GAAAGAGCCTATCAGTTTGTAGTGAACACTTGAGA 825
 QY 928 CCTTATACAAATATCTTTTGAAGTTTCAGCTGCTTACAACTGAAAGCAGTTATATTTGAT 987
 DB 826 CCTTATACAAATATCTTTTGAAGTTTCAGCTGCTTACAACTGAAAGCAGTTATATTTGAT 885
 QY 988 AGTACGATTTGACAAACCAAGATCAGTCTCTGAAGGACCAACCAAAACTGCGTAAACA 1047
 DB 886 AGTACGATTTGACAAACCAAGATCAGTCTCTGAAGGACCAACCAAAACTGCGTAAACA 945
 QY 1048 GGCACATCAGAGGAAGTCTTTTCAATTTTATGGGACCCACCACTATAGTAAACAGG 1107
 DB 946 GGCACATCAGAGGAAGTCTTTTCAATTTTATGGGACCCCACTATAGTAAACAGG 1005
 QY 1108 AAATTTAGTTATAGAGTTGAATTTATATGAGCAT---CAGGTGCGATTTTGGATAACAGC 1164
 DB 1006 AAATTTAGTTATAGAGTTGAATTTATATGAGCATCAGCAGGTGCGATTTTGGATAACAGC 1065
 QY 1165 ACAAAGACCTCAAGTTTGCAATTCATAACCTTAAACCAATTTTCAATGATGATCTAT 1224
 DB 1066 ACAAAGACCTCAAGTTTGCAATTCATAACCTTAAACCAATTTTCAATGATGATCTAT 1125
 QY 1225 ATTGGGCTGAAACCAAGTGCAGGAGCTGGGCCCAAGTCAATATTTTCAATGATGATCTAT 1284
 DB 1126 ATTGGGCTGAAACCAAGTGCAGGAGCTGGGCCCAAGTCAATATTTTCAATGATGATCTAT 1185
 QY 1285 CCAGATGTTTCAGGGGCGAGTGTTCATTTTCAACTTTGAGAGGTAGAATTCACGCAAGTA 1344
 DB 1186 CCAGATGTTTCAGGGGCGAGTGTTCATTTTCAACTTTGAGAGGTAGAATTCACGCAAGTA 1245
 QY 1345 AGAATTTACTTGGAGAAACCAAGCAACCAATGGAATTTTAAACCAATACCGAGTGA 1404
 DB 1246 AGAATTTACTTGGAGAAACCAAGCAACCAATGGAATTTTAAACCAATACCGAGTGA 1305
 QY 1405 GTGCTAGTTCCAGAGACAGGAATATTTTGGAAATATCTTTGCTCCTCCTGGAATATAG 1464
 DB 1306 GTGCTAGTTCCAGAGACAGGAATATTTTGGAAATATCTTTGCTCCTCCTGGAATATAG 1365
 QY 1465 TATATAAATGACCCCATGCTCCAGAAATTTGGAACATAGTAGAGCCCAATGTTAGATTA 1524
 DB 1366 ---ATPAATGACCCCATGCTCCAGAAATTTGGAACATAGTAGTAGAGCCCAATGTTAGATTA 1422
 QY 1525 TATGAGGGTTCAGCAGAGATGTCGTGACCTTCACTCACTTGTGTACATTTTATATATAAC 1584

Db 1423 TATGAGGTTACGAGAGATGCTGACCTTCACTCACTGCTACATTTATATAAC 1482
QY 1585 AGCCATCCAGATAAAACCTTCTGCAAGAAATAGAGCTGGAAGACAGACTTCCACAGTT 1644
Db 1483 AGCCATCCAGATAAAACCTTCTGCAAGAAATAGAGCTGGAAGACAGACTTCCACAGTT 1542
QY 1645 GTAACTACAGGAATCAGTATATTAAGTATGATGAGCTGGAAGCTGCTTATGTTATC 1704
Db 1543 GTAACTACAGGAATCAGTATATTAAGTATGATGAGCTGGAAGCTGCTTATGTTATC 1602
QY 1705 AGGAGACTTGTACCTTTCACAGCACATGATTTAGTGTATCTGCTTTCACCAATCATGGGA 1764
Db 1603 ATCAGATTAAGGAGATTTGGGCTGAGACAATGGGGTTTCTAGATATACATCATG --- 1659
QY 1765 GAAGGACCAACCAAGTCTCAGGTGTAGGACACGTCAGCAAGTGCGCAAGCTCCATTA 1824
Db 1660 -----TCATCTGCAAGCAGGACAAATTTGACTTCCCGAGGCCCTTTGTGAGCCCAA 1710
QY 1825 ATTATAAACTATAAAATATTAGTTCTTCATCTATTTTGTGTATATTGGGATCCTCAGAA 1884
Db 1711 AATTTACAGATTACATGTTTACCATAACAGAGTATTTTACACTGGGATCCTCAGAT 1770
QY 1885 TATCCCAATGGAATAAATCACTCACTATACGATTTATGCAATGGAATTTGGATACAAACAGA 1944
Db 1771 CCTGTATTTTTCATCATTTACCTTATCACTATTTGGATGTTGMAAACCAATCCAAGAT 1830
QY 1945 GCATCCAGATACCTACATAGATACAGCTTCTCATACAGGGTTTAAAGAAATACACA 2004
Db 1831 ATTATTTTAAAGGACATTAACAGTTTGTCTCTTGTCTTATAGGGTTTAAAGAAATACACA 1890
QY 2005 AATACAAATGAGAGTGAGCGCTCAACCCAGATGGAAGTCTTGTCTGAGAA 2064
Db 1891 AATACAAATGAGAGTGAGCGCTCAACCCAGTGGGAGAAAGTTCTTGTCTGAGAA 1950
QY 2065 AATGACATCTTGTGAGAACTTCAGAAAGATGAACCGGAATCATCACTCAAGATGTCGAA 2124
Db 1951 AATGACATCTTGTGAGAACTTCAGAAAGATGAACCGGAATCATCACTCAAGATGTCGAA 2010
QY 2125 GTAAATGATTTACCGCAGATGAATTAAGGTTGAAGTGTCACCAACCGAAAGGCCAAT 2184
Db 2011 GTAAATGATTTACCGCAGATGAATTAAGGTTGAAGTGTCACCAACCGAAAGGCCAAT 2070
QY 2185 GGGATCATTTATGCTTATCAAGTGCTATATAAATATAGATACTTTATATAGAGAC 2244
Db 2071 GGGATCATTTATGCTTATCAAGTGCTATATAAATATAGATACTTTATATAGAGAAC 2130
QY 2245 ACATCAACCAACAGACATAATATTAAGGAACTTAAGACCTCAACACCTCTATACATTTCT 2304
Db 2131 ACATCAACCAACAGACATAATATTAAGGAACTTAAGACCTCAACACCTCTATACATTTCT 2190
QY 2305 GTAAGTCTTACACAGATTTGGTCAAGGAAATCAGGATCTTCTTATCTCTGTAAGG 2364
Db 2191 GTAAGTCTTACACAGATTTGGTCAAGGAAATCAGGATCTTCTTATCTCTGTAAGG 2250
QY 2365 ACTTCGGAGCTGCTGCTGATAGTGCACCAAGAAATATCACTTACAAAATATTTCTCT 2424
Db 2251 ACTTCGGAGCTGCTGCTGATAGTGCACCAAGAAATATCACTTACAAAATATTTCTCT 2310
QY 2425 GGAGAGATTTGAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAATCATATAAAATAT 2484
Db 2311 GGAGAGATTTGAGCTATCATTTCTTCCCAAGTAGTCCCAATGGAATCATATAAAATAT 2370
QY 2485 ACAATTTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATAAATCAACCTCTTTA 2544
Db 2371 ACAATTTATCTCAAGAGAGTAAATGGAATGAGGAAGAACTATAAATCAACCTCTTTA 2430
QY 2545 ACCCAAAACATTAAGCTACTGAGAAATATACCAATATATCATTCAGGTGCTGCTAGT 2604
Db 2431 ACCCAAAACATTAAGGTCTGAGAAATATATACCAATATATCATTCAGGTGCTGCTAGT 2490
QY 2605 ACATGAAAGGTGAAGGAGTTCGGAGTCTCCCATTAAGTATATCTGACGGGAGGATGCT 2664

Db 2491 ACACCTCAAAAGGTGAAGGAGTTGCGAGTGTCTCCCAATAAGTATATCTGACGGGAGGAGTGCT 2550
QY 2665 CCGTATTTCTCCCCCTCAAGACTTCTCTGTAAAAACAGTTGTCTGGTGTCAACGGTGAAGTTG 2724
Db 2551 CCGTATTTCTCCCCCTCAAGACTTCTCTGTAAAAACAGTTGTCTGGTGTCAACGGTGAAGTTG 2610
QY 2725 TCATGGCAACCAACCCCTCGAGCCAAATGGAATTTATCTTTATTAACAAGTTTATGCTCTGG 2784
Db 2611 TCATGGCAACCAACCCCTCGAGCCAAATGGAATTTATCTTTATTAACAAGTTTATGCTCTGG 2670
QY 2785 AATAGATCATCATTTAAAAAATATTAAATGTCACTGAAACATCATTCGGAGTTATCAGATTG 2844
Db 2671 ---AGATCATCATTTAAAAAATATTAAATGTCACTGAAACATCATTCGGAGTTATCAGATTG 2727
QY 2845 GATTTAATGTTGAATACAGTCTTATGTAAACAGCTAGCACCAAGATTTGGTGTATGGGAA 2904
Db 2728 GATTTAATGTTGAATACAGTCTTATGTAAACAGCTAGCACCAAGATTTGGTGTATGGGAA 2787
QY 2905 ACAGGAAGCAATATATCTTAGCTTTCAAAACACAGAGGGAGCACCAAGCGATCCTCCCAA 2964
Db 2788 ACAGGAAGCAATATCTTAGCTTTCAAAACACAGAGGG---ACCAAGCGATCCTCCCAA 2844
QY 2965 GATGTTTATTAAGCAACCTCAGTTCTTCAATCAATTTCTTTTCTGGACACCTCTTCA 3024
Db 2845 GATGTTTATTAAGCAACCTCAGTTCTTCAATCAATTTCTTTTCTGGACACCTCTTCA 2904
QY 3025 AAACCTAATGGGATTTACAAATATTTACTCTGTTTATTAACAGAAATATCTTCAGGTACTTTT 3084
Db 2905 AAACCTAATGGGATTTACAAATATTTACTCTGTTTATTAACAGAAATATCTTCAGGTACTTTT 2964
QY 3085 ATGCAAGATTTTACACTCCTCATGAACCTAACCAATGACTTTGCAATATGACTGTATCCACA 3144
Db 2965 ATGCAAGATTTTACACTCCTCATGAAGTAAACCAATGACTTTGCAATATGACTGTATCCACA 3024
QY 3145 ATTATAGATAAATGCAATATTTAGCTACTATATACATTTGGTTTAAACAGCAAGTACTTCA 3204
Db 3025 ATTTAGATATAACTGCAATATTTAGCTACTATATACATTTGGTTTAAACAGCAAGTACTTCA 3084
QY 3205 GTTGAATGGGAATTAAGAGCTGACATCATTTGAAGTATATACAGATCAAGACATCT 3264
Db 3085 GTTGAATGGGAATTAAGAGCTGACATCATTTGAAGTATATACAGATCAAGACATCT 3144
QY 3265 GAAGGTTTGTGGAAACCTGACTTACGAATCCATTTGCTCAATGCAATAAATGTGAAGC 3324
Db 3145 GAAGGTTTGTGGAAACCTGACTTACGAATCCATTTGCTCAATGCAATAAATGTGAAGC 3204
QY 3325 TGGGTTCCACCGGCTCAACCAACCGGCTAGTCTTCTACTATGTTTCACTGATCTTACAG 3384
Db 3205 TGGGTTCCACCGGCTCAACCAACCGGCTAGTCTTCTACTATGTTTCACTGATCTTACAG 3264
QY 3385 CAGACTCTCGCATGTGAGCCACCTCTTGTATATATGAGAGAGCATATATTTTGTAT 3444
Db 3265 CAGACTCTCGCATGTGAGCCACCTCTTGTATATATGAGAGAGCATATATTTTGTAT 3324
QY 3445 AATCTGAAAAATATACCTGATTTATATAATAAAAAATTTAGTATTAACCTTAAAGGAGGATTC 3504
Db 3325 AATCTGAAAAATATACCTGATTTATATAATAAAAAATTTAGTATTAACCTTAAAGGAGGATTC 3384
QY 3505 TCTGATACCTATATGTCGCCAGCTATACATCAAGACTGAAGAGAGTGTCCAGAAATCTTCA 3564
Db 3385 TCTGATACCTATATGTCGCCAGCTATACATCAAGACTGAAGAGAGTGTCCAGAAATCTTCA 3444
QY 3565 CCAATTAATCAACCTTTTAAACCTTCTCTACTCTGATCTTCTTATCATGGGATCCC 3624
Db 3445 CCAATTAATCAACCTTTTAAACCTTCTCTACTCTGATCTTCTTATCATGGGATCCC 3504
QY 3625 CCAGTAAAGCCAAATGGTGCAATAAATGATTTATGATTTAACTTTAACTTAAAGGAGCAAAATGAA 3684
Db 3505 CCAGTAAAGCCAAATGGTGCAATAAATGATTTATGATTTAACTTTAACTTAAAGGAGCAAAATGAA 3564
QY 3685 AATATTCTTTCACTTCTGATTAATTTACATAAATATTTGGAAGAGCTTTTACATTTTACA 3744
Db 3565 AATATTCTTTCACTTCTGATTAATTTACATAAATATTTGGAAGAGCTTTTACATTTTACA 3624

QY 3745 TTATATAGCTTTTGTGTCGCCGCAAGAACTAGAAAAAGGACTTGGTCTTCCAGTATTCCT 3804
DB 3625 TTATATAGCTTTTGTGTCGCCGCAAGAACTAGAAAAAGGACTTGGTCTTCCAGTATTCCT 3684
QY 3805 TTCTTTTACACAGATCAGTCCAGTCCGCTTACGACCTCCACAAATTTGACTTTAAATCAAC 3864
DB 3685 TTCTTTTACACAGATCAGTCCAGTCCGCTTACGACCTCCACAAATTTGACTTTAAATCAAC 3744
QY 3865 TGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGCTATGTT 3924
DB 3745 TGTACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGCTATGTT 3804
QY 3925 AAGTATATAGTTTAAAAATTCATGAACATGAACCTGACACTATATATTAAGAATATA 3984
DB 3805 AAGTATATAGTTTAAAAATTCATGAACATGAACCTGACACTATATATTAAGAATATA 3864
QY 3985 TCAGGAATTTAAACTGAAGCAAACTTGTGGACTGGAACCCAGTCCAGTCTACTCTATC 4044
DB 3865 TCAGGAATTTAAACTGAAGCAAACTTGTGGACTGGAACCCAGTCCAGTCTACTCTATC 3924
QY 4045 CGTGTATCTCGCTTACCAAGTTGGAATGGCAATCAATTTAGTAAATAGTATTAATTC 4104
DB 3925 CGTGTATCTCGCTTACCAAGTTGGAATGGCAATCAATTTAGTAAATAGTATTAATTC 3984
QY 4105 ACAACCAAGATCAGTCCAGATGTCGTCAGAAATATGCAAGTCATGGCAACTAGCTGG 4164
DB 3985 ACAACCAAGATCAGTCCAGATGTCGTCAGAAATATGCAAGTCATGGCAACTAGCTGG 4044
QY 4165 CAGTCAGTTTATAGTAAATGGATCCACCCAAAGGCAAAATGGGAATTAATGCGAGTAT 4224
DB 4045 CAGTCAGTTTATAGTAAATGGATCCACCCAAAGGCAAAATGGGAATTAATGCGAGTAT 4104
QY 4225 ATGGTAAACAGTTGAAAGGAATTCACAAAAGTTTCTCCCAAGATCAATGTACACTTTC 4284
DB 4105 ATGGTAAACAGTTGAAAGGAATTCACAAAAGTTTCTCCCAAGATCAATGTACACTTTC 4164
QY 4285 ATAAAGCTTCTTGCCAAATACCTCATATGTTCTTTAAAGTAAAGCTTCAACCTCAGCTGT 4344
DB 4165 ATAAAGCTTCTTGCCAAATACCTCATATGTTCTTTAAAGTAAAGCTTCAACCTCAGCTGT 4224
QY 4345 GAAGGTGATGAAGCACTGCCATGTCAGACACTACCTGAAACAGTTCCTCAGTGTTCCT 4404
DB 4225 GAAGGTGATGAAGCACTGCCATGTCAGACACTACCTGAAACAGTTCCTCAGTGTTCCT 4284
QY 4405 ACAATATTTGCTTTTCTGATGTTCACTCACTAGTGCAACATTTGATGGATTAAGACCT 4464
DB 4285 ACAATATTTGCTTTTCTGATGTTCACTCACTAGTGCAACATTTGATGGATTAAGACCT 4344
QY 4465 GACACTATCTTGGCTACTTTCAAAATTAACAAATTTACCACTCAACTCTGCTCAAAA 4524
DB 4345 GACACTATCTTGGCTACTTTCAAAATTAACAAATTTACCACTCAACTCTGCTCAAAA 4404
QY 4525 TGCAGAAGATGGGAATCCGAAGATGTTGTAATATCAAAAAATTCATACCTCTATGAA 4584
DB 4405 TGCAGAAGATGGGAATCCGAAGATGTTGTAATATCAAAAAATTCATACCTCTATGAA 4464
QY 4585 GCTCACTTAACTGAAGACAGTATATGATTAAGAAATTTAGATGGTATAGATTCCAA 4644
DB 4465 GCTCACTTAACTGAAGACAGTATATGATTAAGAAATTTAGATGGTATAGATTCCAA 4524
QY 4645 GTGGCTGCCAGCACCAATGCTGCTATGGAATGCTTCAAACTGGATTTCTACAAAACT 4704
DB 4525 GTGGCTGCCAGCACCAATGCTGCTATGGAATGCTTCAAACTGGATTTCTACAAAACT 4584
QY 4705 GTGGCTGGCCCTCCAGATGGTCTCTCGAAATATGTTTCATGTAGTAGCAACATCACTTTT 4764
DB 4585 GTGGCTGGCCCTCCAGATGGTCTCTCGAAATATGTTTCATGTAGTAGCAACATCACTTTT 4644
QY 4765 AGCATCAGCATAGCTGGAGTGAACCTGCTGCTATTTACCTGGAACCATGTTATCTGATT 4824
DB 4645 AGCATCAGCATAGCTGGAGTGAACCTGCTGCTATTTACCTGGAACCATGTTATCTGATT 4704

QY 4825 GATGTCAAATCGGTAGATTAATGAAATTAATATATATCTCTCATCAAGTCAAAATGAAGAA 4884
DB 4705 GATGTCAAATCGGTAGATTAATGAAATTAATATATATCTCTCATCAAGTCAAAATGAAGAA 4764
QY 4885 AATAAAACCATAGAAATTAAGATTTAGAAATATTACAAAGGTATTCTGTAGTGTACCT 4944
DB 4765 AATAAAACCATAGAAATTAAGATTTAGAAATATTACAAAGGTATTCTGTAGTGTACCT 4824
QY 4945 GCATTTACTGGGAACATTAAGTCTGCTCATATGTAGAGGGAAGTCAAGTCTGGAATGATT 5004
DB 4825 GCATTTACTGGGAACATTAAGTCTGCTCATATGTAGAGGGAAGTCAAGTCTGGAATGATT 4884
QY 5005 GTTACTATCTTTAGATCAGCCCAAGAGGACCACTTAACACATGACATTTGAGAAGATA 5064
DB 4885 GTTACTATCTTTAGATCAGCCCAAGAGGACCACTTAACACATGACATTTGAGAAGATA 4944
QY 5065 CCAGATGAAGTTTACAAAATTTCAATTTAAGCTTCTTCTCTCAACTCAATGAAAT 5124
DB 4945 CCAGATGAAGTTTACAAAATTTCAATTTAAGCTTCTTCTCTCAACTCAATGAAAT 5004
QY 5125 ATCCAAGTATATCAAGCTCTGCTTTTACCGAGAAGATGATCTTCTGCTGCTCCAGATTCCAC 5184
DB 5005 ATCCAAGTATATCAAGCTCTGCTTTTACCGAGAAGATGATCTTCTGCTGCTCCAGATTCCAC 5064
QY 5185 AACCTCAGTATTAATACAGAAAACCAACACATTCGTCATTTGCAATGCTAGAGGACTAAAA 5244
DB 5065 AACCTCAGTATTAATACAGAAAACCAACACATTCGTCATTTGCAATGCTAGAGGACTAAAA 5124
QY 5245 GGTGACATACATACATATCAGTGTTTACGCAAGTCAATAGTGTGCTGCTGCTCCAAAG 5304
DB 5125 GGTGACATACATACATATCAGTGTTTACGCAAGTCAATAGTGTGCTGCTGCTCCAAAG 5184
QY 5305 GTTCGATGAGAAATTAACCATGATATCAAGCTCCAGCAACGACCAAAAAACCAACCAACC 5364
DB 5185 GTTCGATGAGAAATTAACCATGATATCAAGCTCCAGCAACGACCAAAAAACCAACCAACC 5244
QY 5365 CCTATTATGATGCAAGGAAAACTGCTTTGTCATTTCAACAAATTAATCAATCAAGATG 5424
DB 5245 CCTATTATGATGCAAGGAAAACTGCTTTGTCATTTCAACAAATTAATCAATCAAGATG 5304
QY 5425 CCAATATGTTACTACAGTGTATCATGGACCAATTAATAAATGTAAGTGTGCTGCTGCA 5484
DB 5305 CCAATATGTTACTACAGTGTATCATGGACCAATTAATAAATGTAAGTGTGCTGCTGCA 5364
QY 5485 GAAACAGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGTCATATTTAATAAA 5544
DB 5365 GAAACAGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGTCATATTTAATAAA 5424
QY 5545 GCAAGGCAATTTTAAATGAAGGCTTCTTAAACCTCCATGTAAGAGGAAAGACA 5604
DB 5425 GCAAGGCAATTTTAAATGAAGGCTTCTTAAACCTCCATGTAAGAGGAAAGACA 5484
QY 5605 AAGTTTGTGGCAATGAAGAAATCTACATCATAGTGTGTAATGTCATGATGCTCT 5664
DB 5485 AAGTTTGTGGCAATGAAGAAATCTACATCATAGTGTGTAATGTCATGATGCTCT 5544
QY 5665 GGCATTAAGACAAAAATTTGCAATGGAACCACTGGAACCAAAAAAGCAATCTTATTAAAA 5724
DB 5545 GGCATTAAGACAAAAATTTGCAATGGAACCACTGGAACCAAAAAAGCAATCTTATTAAAA 5604
QY 5725 TTTAGAGCTACAAATTAATATGGGAACAATTTACTGCTCTGATTTCTGACCTGCTGTAAG 5784
DB 5605 TTTAGAGCTACAAATTAATATGGGAACAATTTACTGCTCTGATTTCTGACCTGCTGTAAG 5664
QY 5785 ACTTTAGGGGAGGACTTTCAGAAAGAACCGTAGAGATCATTCTTCTCGTCACTTTGCT 5844
DB 5665 ACTTTAGGGGAGGACTTTCAGAAAGAACCGTAGAGATCATTCTTCTCGTCACTTTGCT 5724
QY 5845 ATCTCTTCAATTAATCTCTCTGGAACAGCTATTTTGTGATTTTGAAGATTTGCAAGATTCGACAGAG 5904
DB 5725 ATCTCTTCAATTAATCTCTCTGGAACAGCTATTTTGTGATTTTGAAGATTTGCAAGAG 5784
QY 5905 CAGAAAGAGGTGGCACTACTCTCTCAGGATGCAAGAAATTTATTGACACTAAATTTGAAG 5964

Db 64 CATCATCAAGACCAAAAGTAGATAAAAACACAAGATGGGAAAAAAGACAGACGAAAA 123
QY 1249 ACTGGGCCCAAGTCAAAATATTTTCAGTATTCACTCCACAGATGTTCCAGGGGAGTGT 1308
Db 124 ACTGGAACTCTAAAAAGACAGACGACCTCTCTCTCTCCAAAGGATGGCAGGGCAGTGT 183
QY 1309 GATTTACAACTTCGAGGGTAGAATCCAGCAAGTAAAGTAATCTTGGAGAAACACCA 1368
Db 184 GATTTACAACTTCGAGGGTAGAATCCAGCAAGTAAAGTAATCTTGGAGAAACACCA 243
QY 1369 CAACCAATGGAATTTATTAACCAATACCGAGTGAAGTGTCTTCCAGAGACAGGAATA 1428
Db 244 CAACCAATGGAATTTATTAACCAATACCGAGTGAAGTGTCTTCCAGAGACAGGAATA 303
QY 1429 ATTTGGAAATATCTTTGCTCACTGGAAATAATGAGTATATAAATGACCCATCGGTCCA 1488
Db 304 ATTTGGAAATATCTTTGCTCACTGGAAATAATGAGTATATAAATGACCCATCGGTCCA 363
QY 1489 GAAATTTGNAACATAGTAGGCCAATGGTAGGATTTATGAGGGTTTCAGCAGAGATGTCG 1548
Db 364 GAAATTTGNAACATAGTAGGCCAATGGTAGGATTTATGAGGGTTTCAGCAGAGATGTCG 423
QY 1549 TCTGACCTTCACCTTCCTGCTACATTTATATAAAGCCATCCAGATCAAAACTTTTCT 1608
Db 424 TCTGACCTTCACCTTCCTGCTACATTTATATAAAGCCATCCAGATCAAAACTTTTCT 483
QY 1609 GCAAGGAATAGAGCTGAAGACAGACTTCAACAGTTGTAACTAAGGAATCAGTATAT 1668
Db 484 GCAAGGAATAGAGCTGAAGACAGACTTCAACAGTTGTAACTAAGGAATCAGTATAT 543
QY 1669 ACTGACATTCAGCTGAAAGCTGTCTTATGTTATCAGGAGCTTGTACCTTCAGTGA 1728
Db 544 ACTGACATTCAGCTGAAAGCTGTCTTATGTTATCAGGAGCTTGTACCTTCAGTGA 603
QY 1729 CACATGATTAGTGTATCTGCTTTCACATCATCGGAGAGGACCAACAAGTTCTCAGT 1788
Db 604 CACATGATTAGTGTATCTGCTTTCACATCATCGGAGAGGACCAACAAGTTCTCAGT 663
QY 1789 GTTAGGACAGCTCAGCAAGTGCACAGCTCCATTTAAATATATAAATATAAATATTAGT 1848
Db 664 GTTAGGACAGCTCAGCAAGTGCACAGCTCCATTTAAATATATAAATATAAATATTAGT 723
QY 1849 TCTTCATCTATTTTGTATATTTGGATCTCTCAGATATATCCCAATGGAAAAATACTCAC 1908
Db 724 TCTTCATCTATTTTGTATATTTGGATCTCTCAGATATATCCCAATGGAAAAATACTCAC 783
QY 1909 TATACGATTTATGCAATGGAATTTGGATACAAACAGAGCATTCAGATTAACCTACATAGAT 1968
Db 784 TATACGATTTATGCAATGGAATTTGGATACAAACAGAGCATTCAGATTAACCTACATAGAT 843
QY 1969 AACAGCTTTTCTATAACAGGGTTTAAAGAAATACAAAAATGAGAGTGGCAGCC 2028
Db 844 AACAGCTTTTCTATAACAGGGTTTAAAGAAATACAAAAATGAGAGTGGCAGCC 903
QY 2029 TCAACCCAGATGGAGAAAGTTCTTGTCTGAAAGAAATGACATCTTTGTGAGAACTTCA 2088
Db 904 TCAACCCAGATGGAGAAAGTTCTTGTCTGAAAGAAATGACATCTTTGTGAGAACTTCA 963
QY 2089 GAGATGAAACCGGAATCATCACTCAAGATGTCGAGTAAATGATGTTACCCAGATGAA 2148
Db 964 GAGATGAAACCGGAATCATCACTCAAGATGTCGAGTAAATGATGTTACCCAGATGAA 1023
QY 2149 ATAAAGTTGAAGTGGTCACCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGAAGTG 2208
Db 1024 ATAAAGTTGAAGTGGTCACCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGAAGTG 1083
QY 2209 CTATATAAATAATAGATCTTTTATATGAAGAACACATCAACCAAGATATATTA 2268
Db 1084 CTATATAAATAATAGATCTTTTATATGAAGAACACATCAACCAAGATATATTA 1143
QY 2269 AGGAACTTAAGACCTCACACCTCTATTAATCTTGTAAAGTCTTACACGATTTGGT 2328
Db 1144 AGGAACTTAAGACCTCACACCTCTATTAATCTTGTAAAGTCTTACACGATTTGGT 1203

QY 2329 CATGGCAATCAGGTATCTTCTTTTACTCTCTGTAAAGACTTTCGGAGACTGTGCTGTATGT 2388
Db 1204 CATGGCAATCAGGTATCTTCTTTTACTCTCTGTAAAGACTTTCGGAGACTGTGCTGTATGT 1263
QY 2389 GCACAGAAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCCCT 2448
Db 1264 GCACAGAAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTTCCCT 1323
QY 2449 CCCCCAAGTAGTCCCAATGGAATCAAAAAAATATACAATTTATCTCAAGAGAACTAAT 2508
Db 1324 CCCCCAAGTAGTCCCAATGGAATCAAAAAAATATACAATTTATCTCAAGAGAACTAAT 1383
QY 2509 GGAATAGGAAAGAACTATAAATACAACTCTTTTAAACCAAAACATTAAGTACTGAAG 2568
Db 1384 GGAATAGGAAAGAACTATAAATACAACTCTTTTAAACCAAAACATTAAGTACTGAAG 1443
QY 2569 AAATATACCCCAATATATCAATGAGTGTCTGTAGTACACTGAAAGGTGAAGAGTTCCG 2628
Db 1444 AAATATACCCCAATATATCAATGAGTGTCTGTAGTACACTGAAAGGTGAAGAGTTCCG 1503
QY 2629 AGTGTCTCCCATAGTATACTGACGGAAGAGTGTCTCTGTATCTCCCTCAAGACTTC 2688
Db 1504 AGTGTCTCCCATAGTATACTGACGGAAGAGTGTCTCTGTATCTCCCTCAAGACTTC 1563
QY 2689 TCTGTAAAAACAGTTGTCTGTGTACGGTGAAGTGTCTATGAGCAACCAACCTTGGAGCCA 2748
Db 1564 TCTGTAAAAACAGTTGTCTGTGTACGGTGAAGTGTCTATGAGCAACCAACCTTGGAGCCA 1623
QY 2749 AATGGAATTTATCTTTTATACACAGTTTATGCTGGAATAGATCATTTAAAACTATT 2808
Db 1624 AATGGAATTTATCTTTTATACACAGTTTATGCTGGAATAGATCATTTAAAACTATT 1683
QY 2809 AATGCACTGAAACATCATTTGGAGTTATCAGATTTTGGATTTAAATGTTGAATACAGTGT 2868
Db 1684 AATGCACTGAAACATCATTTGGAGTTATCAGATTTTGGATTTAAATGTTGAATACAGTGT 1743
QY 2869 ATGTAAACAGCTAGACCAAGATTTGGTGAAGGAAAAAGGAAGCAATATCATTAGCTTT 2928
Db 1744 ATGTAAACAGCTAGACCAAGATTTGGTGAAGGAAAAAGGAAGCAATATCATTAGCTTT 1803
QY 2929 CAAACACAGAGGGGACCAAGCCATCTCCCAAGATGTTTATATGCAACACCTCAGT 2988
Db 1804 CAAACACAGAGGGGACCAAGCCATCTCCCAAGATGTTTATATGCAACACCTCAGT 1863
QY 2989 TCTTCATCAATTAATCTTTTCTGGACACCTCTCCCAAGATGTTTATATGCAACACCTCAGT 3048
Db 1864 TCTTCATCAATTAATCTTTTCTGGACACCTCTCCCAAGATGTTTATATGCAACATAT 1923
QY 3049 TACTCTGTTTATTAAGAAATACTTCAGGTACTTTTATGCAAGATTTTACATCCATGAA 3108
Db 1924 TACTCTGTTTATTAAGAAATACTTCAGGTACTTTTATGCAAGATTTTACATCCATGAA 1983
QY 3109 CTAAACCAATGATTTGACAAATGATGTTATCCCAATTTATAGTAACTGCAATATTC 3168
Db 1984 GTAAACCAATGATTTGACAAATGATGTTATCCCAATTTATAGTAACTGCAATATTC 2043
QY 3169 AGCTACTATACATTTTGGTTTAAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGT 3228
Db 2044 AGCTACTATACATTTTGGTTTAAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAGCAGT 2103
QY 3229 GACATCATTTGAAGTATACACAGATCAAGATCACTTGAAGGGTTCCTGGAAACCTGACT 3288
Db 2104 GACATCATTTGAAGTATACACAGATCAAGATCACTTGAAGGGTTCCTGGAAACCTGACT 2163
QY 3289 TAGGAATCAATTTGCTCAACTGCAATTAATGTAAGCTGGGTCCACCGGCTCAACCAAC 3348
Db 2164 TAGGAATCAATTTGCTCAACTGCAATTAATGTAAGCTGGGTCCACCGGCTCAACCAAC 2223
QY 3349 GGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCTCGGCACTGTGAGACCA 3408
Db 2224 GGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTCTCTCGGCACTGTGAGACCA 2283

QY 3409 CCTCTGTGTACATATGAGAGAGCATATATTTTGTAAATCTGGAAAAATACACTGATTTAT 3468
DB |||||
QY 2284 CCTCTGTGTACATATGAGAGAGCATATATTTTGTAAATCTGGAAAAATACACTGATTTAT 2343
DB |||||
QY 3469 ATATTAAAAATTTACTCCATCAACAGAAAAAGGATTTCTCTGATACCTTATCTGCCCCAGCTA 3528
DB |||||
QY 2344 ATATTAAAAATTTACTCCATCAACAGAAAAAGGATTTCTCTGATACCTTATCTGCCCCAGCTA 2403
DB |||||
QY 3529 TACATCAAGCTGAAGAGATGTCACAGAAATCTTCAACCAATTAATCAACACTTTTAAAAAC 3588
DB |||||
QY 2404 TACATCAAGACTGAAGAGATGTCACAGAAATCTTCAACCAATTAATCAACACTTTTAAAAAC 2463
DB |||||
QY 3589 CTTTCTCTACTCTCAGTCTCTTATCATGCGATCCCCAGTAAAGCCAAATGGTCAATA 3648
DB |||||
QY 2464 CTTTCTCTACTCTCAGTCTCTTATCATGCGATCCCCAGTAAAGCCAAATGGTCAATA 2523
DB |||||
QY 3649 ATAAGTTATGATTTAACTTTTACAAGGACCAAAATGAAAAATTAATCTTTTCAATTTCTGAT 3708
DB |||||
QY 2524 ATAAGTTATGATTTAACTTTTACAAGGACCAAAATGAAAAATTAATCTTTTCAATTTCTGAT 2583
DB |||||
QY 3709 AATTACATATATTTGAAGAGCTTTTCAACATTTTACATTTATATAGCTTTTTCGTCGCCA 3768
DB |||||
QY 2584 AATTACATATATTTGAAGAGCTTTTCAACATTTTACATTTATATAGCTTTTTCGTCGCCA 2643
DB |||||
QY 3769 AGAACTAGAAAAAGGACTTGGTCTTCCAGTATTTCTTTTCAACAGATGAGTCAGTG 3828
DB |||||
QY 2644 AGNACTAGAAAAAGGACTTGGTCTTCCAGTATTTCTTTTCAACAGATGAGTCAGTG 2703
DB |||||
QY 3829 CGGTTAGCACTTCCACAAAAATTTGACTTTTAACTGACTTCACTTCAAGCTTTGTATGGCTG 3888
DB |||||
QY 2704 CGGTTAGCACTTCCACAAAAATTTGACTTTTAACTGACTTCACTTCAAGCTTTGTATGGCTG 2763
DB |||||
QY 3889 AATGAGGCCCAAGTCTCTTCCAGTGTATTTGTTTAAAGTATATAGTTTAAAAATTCAT 3948
DB |||||
QY 2764 AATGAGGCCCAAGTCTCTTCCAGTGTATTTGTTTAAAGTATATAGTTTAAAAATTCAT 2823
DB |||||
QY 3949 GAACTAGAACTGACACTATATATTAAGAAATATATCAGGATTTAAAACTGAAGCCAAA 4008
DB |||||
QY 2824 GAACTAGAACTGACACTATATATTAAGAAATATATCAGGATTTAAAACTGAAGCCAAA 2883
DB |||||
QY 4009 CTTGTTGGACTGGAACCACTGACACCTACTCTATCCGTGTATCTGCGTTTCAACCAAGTT 4068
DB |||||
QY 2884 CTTGTTGGACTGGAACCACTGACACCTACTCTATCCGTGTATCTGCGTTTCAACCAAGTT 2943
DB |||||
QY 4069 GGAATGGCAATCAATTTAGTAATGTAGTAATTTCAACCAAGAACTCAGTTCCAGAT 4128
DB |||||
QY 2944 GGAATGGCAATCAATTTAGTAATGTAGTAATTTCAACCAAGAACTCAGTTCCAGAT 3003
DB |||||
QY 4129 GTCTGCGAATATGCAATGCACTGCGCAACTAGCTGGCAGTCAGTTTGTAGTAAATGGAT 4188
DB |||||
QY 3004 GTCTGCGAATATGCAATGCACTGCGCAACTAGCTGGCAGTCAGTTTGTAGTAAATGGAT 3063
DB |||||
QY 4189 CCAACCAAAAGGCAAAATGGAATTAATACGCAATATATGTTAAGCAAGTTCGAAAGAAATTC 4248
DB |||||
QY 3064 CCAACCAAAAGGCAAAATGGAATTAATACGCAATATATGTTAAGCAAGTTCGAAAGAAATTC 3123
DB |||||
QY 4249 ACBAAGTTTCTCCCAAGATCAGATGACATTTTCAATTAAGCTTTTCCCAATACCTTCA 4308
DB |||||
QY 3124 ACBAAGTTTCTCCCAAGATCAGATGACATTTTCAATTAAGCTTTTCCCAATACCTTCA 3183
DB |||||
QY 4309 TATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAGAGCATGCCAT 4368
DB |||||
QY 3184 TATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAGAGCATGCCAT 3243
DB |||||
QY 4369 GTCAGACACTACTCTGAAAACAGTTCCTCAGTGTTCACAAAAATATTTGCTTTTCTGATGTT 4428
DB |||||
QY 3244 GTCAGACACTACTCTGAAAACAGTTCCTCAGTGTTCCTCAGTGTTCCTCAGTGTTCCTGATGTT 3303
DB |||||
QY 4429 CAGTCAACTAGTGCAACTGATGATGATGAAGACCTGACACTATCTCTGGCTACTTTCAA 4488
DB |||||
QY 3304 CAGTCAACTAGTGCAACTGATGATGATGAAGACCTGACACTATCTCTGGCTACTTTCAA 3363
DB |||||
QY 4489 AATTACAAAAATTACCCTCAACTCTGCTGCTCAAAAATGCAAGAATGGGAATCCGAGAA 4548

DB |||||
QY 3364 AATTACAAAAATTACCCTCAACTCAACTCTGCTGCTCAAAAATGCAAGAATGGAAATCCGAGAA 3423
DB |||||
QY 4549 TGTGTTGAATATCAAAAAATTCAAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTA 4608
DB |||||
QY 3424 TGTGTTGAATATCAAAAAATTCAAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTA 3483
DB |||||
QY 4609 TATGATTTAAAGAAATTTAGATGGTATAGATTTCCAAAGTGGCTGCCAGACCAATGCTGCG 4668
DB |||||
QY 3484 TATGATTTAAAGAAATTTAGATGGTATAGATTTCCAAAGTGGCTGCCAGACCAATGCTGCG 3543
DB |||||
QY 4669 TATGCAATTTGCTTTCAAACTGGATTTCTACAAAAATCTGCTGCGCTCCAGATGGTCT 4728
DB |||||
QY 3544 TATGCAATTTGCTTTCAAACTGGATTTCTACAAAAATCTGCTGCGCTCCAGATGGTCT 3603
DB |||||
QY 4729 CTTGAAAAATTTGCTATGATGATAGCAATCACTTTTAGCATCAGATGAAGCTGGAGTGA 4788
DB |||||
QY 3604 CTTGAAAAATTTGCTATGATGATAGCAATCACTTTTAGCATCAGATGAAGCTGGAGTGA 3663
DB |||||
QY 4789 CCTGCTGTCTTACTTGGACCAACATGTTATCTGATTTGATGTCAAAATCGGTAGATATGAT 4848
DB |||||
QY 3664 CCTGCTGTCTTACTTGGACCAACATGTTATCTGATTTGATGTCAAAATCGGTAGATATGAT 3723
DB |||||
QY 4849 GAAATTAATATATCTTCTCATCAAGTCAAAATGAAGAAAAATAAAACCATAGAAATTAAGAT 4908
DB |||||
QY 3724 GAAATTAATATATCTTCTCATCAAGTCAAAATGAAGAAAAATAAAACCATAGAAATTAAGAT 3783
DB |||||
QY 4909 TTAGAAAAATTTCAAAAGGATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 4968
DB |||||
QY 3784 TTAGAAAAATTTCAAAAGGATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 3843
DB |||||
QY 4969 GCATATGTAGAGGAGGAGTCAAGTCTGAAATGATTTGTTTACTTACTTACTTACTTACTTACTT 5028
DB |||||
QY 3844 GCATATGTAGAGGAGGAGTCAAGTCTGAAATGATTTGTTTACTTACTTACTTACTTACTTACTT 3903
DB |||||
QY 5029 AAGGACCCCACTTAAACAACATGATTTTCCAGAGATACCAGATGAAGTTTACAAAAATTTCAA 5088
DB |||||
QY 3904 AAGGACCCCACTTAAACAACATGATTTTCCAGAGATACCAGATGAAGTTTACAAAAATTTCAA 3963
DB |||||
QY 5089 TTAAGTTCT 5148
DB |||||
QY 3964 TTAAGTTCT 4023
DB |||||
QY 5149 TACCAGAGAGATGATCT 5208
DB |||||
QY 4024 TACCAGAGAGATGATCT 4083
DB |||||
QY 5209 AACACATTTCT 5268
DB |||||
QY 4084 AACACATTTCT 4143
DB |||||
QY 5269 GTTTACGCAATCAATAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5328
DB |||||
QY 4144 GTTTACGCAATCAATAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4203
DB |||||
QY 5329 ATCAAGGCTCCAGACGACCAAAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAAC 5388
DB |||||
QY 4204 ATCAAGGCTCCAGACGACCAAAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAAC 4263
DB |||||
QY 5389 CTGCTTGTGACTTTCAACCAACAAATTAACAATCAGAATGCCAATATGTTTACTACATGATGAT 5448
DB |||||
QY 4264 CTGCTTGTGACTTTCAACCAACAAATTAACAATCAGAATGCCAATATGTTTACTACATGATGAT 4323
DB |||||
QY 5449 CATGACCAATTAAAAAATGTACAGTGTCTGCGACAGAAACAGGAGCTCAGCATGATGGA 5508
DB |||||
QY 4324 CATGACCAATTAAAAAATGTACAGTGTCTGCGACAGAAACAGGAGCTCAGCATGATGGA 4383
DB |||||
QY 5509 AATGTAAACAAAGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5568
DB |||||
QY 4384 AATGTAAACAAAGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4443
DB |||||
QY 5569 GCGTTTCTTAACTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5628

Db 4444 GGCTTTCCTAACCCCTCCATGTACAGAGGAAAGACAAAGTTTATGTAGTGCATGAGAAATC 4503
Qy 5629 TACATCATAGGTGCTGATTAATGATGATGATTCCTGGCAATGAAGACAAATTTTCAAT 5688
Db 4504 TACATCATAGGTGCTGATTAATGATGATGATTCCTGGCAATGAAGACAAATTTTCAAT 4563
Qy 5689 GGACCACTGAACCAAAAGCAATACTTATTTAAATTTAGAGCTCAAAATATTAGGGA 5748
Db 4564 GGACCACTGAACCAAAAGCAATACTTATTTAAATTTAGAGCTCAAAATATTAGGGA 4623
Qy 5749 CAATTTACTGACTCTGATTAATTTCTGACCCCTGTTAAGACTTTAGGGGAAGGACTTTTCAGAA 5808
Db 4624 CAATTTACTGACTCTGATTAATTTCTGACCCCTGTTAAGACTTTAGGGGAAGGACTTTTCAGAA 4683
Qy 5809 AGAACCGTAGAGATCAATCTTTCCGTCACATTTGTGTATCTTTTCAATAATTTCTCTTGA 5868
Db 4684 AGAACCGTAGAGATCAATCTTTCCGTCACATTTGTGTATCTTTTCAATAATTTCTCTTGA 4743
Qy 5869 ACAGCTATTTTTCATTTTGCAGAAATTCGACAGAGCAAGAGAGAGGTCGCATACTCT 5928
Db 4744 ACAGCTATTTTTCATTTTGCAGAAATTCGACAGAGCAAGAGAGGTCGCATACTCT 4803
Qy 5929 CCTCAGGATGCAGAAATTTATGACACTAAATTCGAAGCTGGATCAGCTCATCAGTGGCA 5988
Db 4804 CCTCAGGATGCAGAAATTTATGACACTAAATTCGAAGCTGGATCAGCTCATCAGTGGCA 4863
Qy 5989 GACCTGGAACTGAAGGACGAGATTAACGCG-----G 6021
Db 4864 GACCTGGAACTGAAGGACGAGATTAACGCGTTACTTAGTTATAGAAATCCATCAAG 4923
Qy 6022 CCAATAGCAAGAAATCTCTCTGCAACATGTTGAAGAGCTTTGCAACAAACCACTTA 6081
Db 4924 CCAATAGCAAGAAATCTCTCTGCAACATGTTGAAGAGCTTTGCAACAAACCACTTA 4983
Qy 6082 AGTTTTCAGAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTTCTTCACTGAT 6141
Db 4984 AGTTTTCAGAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTTCTTCACTGAT 5043
Qy 6142 GCTGATCTGCTTGGATAGAGCAAAACCGTTTCCCAACATAAACCATAATAAT 6201
Db 5044 GCTGATCTGCTTGGATAGAGCAAAACCGCTTCCCAACATAAACCAT---ATAAT 5100
Qy 6202 AATAACAGAGTAAAGCTGATAGTGAAGCTAGTGGTTCCAGGTTCCGATTAATAATGCC 6261
Db 5101 AATAACAGAGTAAAGCTGATAGTGAAGCTAGTGGTTCCAGGTTCCGATTAATAATGCC 5160
Qy 6262 AGCTATTTCTGGTTATTTATGCTCCAAATGATTTATGCTACTCAAGTCCACTACCA 6321
Db 5161 AGCTATTTCTGGTTATTTATGCTCCAAATGATTTATGCTACTCAAGTCCACTACCA 5220
Qy 6322 GGAAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACAGGGCAAAACATTTAGTAATG 6381
Db 5221 GGAAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACAGGCAAAACATTTAGTAATG 5280
Qy 6382 CTAACACAGTGTGTTGAAAGAGCAAGGATCAGATGCCATCAGTATTTGGCCAGAGGCAAC 6441
Db 5281 CTAACACAGTGTGTTGAAAGAGCAAGGATCAGATGCCATCAGTATTTGGCCAGAGGCAAC 5340
Qy 6442 AAGCCAGTTACTGTCTTTGGAGATATAGTATTAACAGCTAATGAGGATGTTCAATA 6501
Db 5341 AAGCCAGTTACTGTCTTTGGAGATATAGTATTAACAGCTAATGAGGATGTTCAATA 5400
Qy 6502 GATTGACTATCAGGATCTGAAATTTGAAAGCATGGGATTTGATGATCTGTCACAG 6561
Db 5401 GATTGACTATCAGGATCTGAAATTTGAAAGCATGGGATTTGATGATCTGTCACAG 5460
Qy 6562 TGTAATCTTACTCCCTGGCCAGAGCATGGGGTTCTTGAGAAACAGCGCCCTCTTAATTCAC 6621
Db 5461 TGTAATCTTACTCCCTGGCCAGAGCATGGGGTTCTTGAGAAACAGCGCCCTCTTAATTCAC 5520
Qy 6622 TTTGTGAAGTGTGTTGAGCAAGCAGGGCAATGACCAACACTAATGATGTTCACTGC 6681
Db 5521 TTTGTGAAGTGTGTTGAGCAAGCAGGGCAATGACCAACACTAATGATGTTCACTGC 5580

Qy 6682 AGTGCTGGAGTTGGAGAACTGGAGTTTATTTATTTGCTCTGGACCATTTTAAACAAACATATA 6741
Db 5581 AGTGCTGGAGTTGGAGAACTGGAGTTTATTTATTTGCTCTGGACCATTTTAAACAAACATATA 5640
Qy 6742 AATGACCATGATTTTGTGGATATATATGACTAGTCTGAACTGAGAGAGTGAAGAAATG 6801
Db 5641 AATGACCATGATTTTGTGGATATATATGACTAGTCTGAACTGAGAGAGTGAAGAAATG 5700
Qy 6802 TGATGCTGAGAACTGCGCAGCATATATCTTTTACACAGTGCATTTCTGGATCTCTTA 6861
Db 5701 TGATGCTGAGAACTGCGCAGCATATATCTTTTACACAGTGCATTTCTGGATCTCTTA 5760
Qy 6862 TCAATATAGGAGAACTAATCAGCCCATCTGTTTGTAACTATTCAGCATTGAGAAGATG 6921
Db 5761 TCAATATAGGAGAACTAATCAGCCCATCTGTTTGTAACTATTCAGCATTGAGAAGATG 5820
Qy 6922 GACTCTTTGGAGCCCATGGAAGTGAATGTTGAGCTTGAATGGAAGAAACCACTATG 6978
Db 5821 GACTCTTTGGAGCCCATGGAAGTGAATGTTGAGCTTGAATGGAAGAAACCACTATG 5877

RESULT 11

ADH41620

ID ADH41620 standard; DNA; 2739 BP.

XX ADH41620;

XX 25-MAR-2004 (first entry)

XX Novel human nucleic acid NOV15e.

ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;
anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;
nootropic; antiparkinsonian; antiasthmatic; antiinfertility;
cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
Parkinson's disease; asthma; fertility disorder; chromosome mapping;
tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX Homo sapiens.

XX WO2003102159-A2.

XX 11-DEC-2003.

XX 04-JUN-2003; 2003WO-US017573.

XX 04-JUN-2002; 2002US-0385490P.

XX 04-JUN-2002; 2002US-0385615P.

XX 04-JUN-2002; 2002US-0385755P.

XX 05-JUN-2002; 2002US-0386041P.

XX 06-JUN-2002; 2002US-0386355P.

XX 06-JUN-2002; 2002US-0386357P.

XX 06-JUN-2002; 2002US-0386447P.

XX 06-JUN-2002; 2002US-0386459P.

XX 06-JUN-2002; 2002US-0386465P.

XX 06-JUN-2002; 2002US-0386864P.

XX 07-JUN-2002; 2002US-0386701P.

XX 07-JUN-2002; 2002US-0386796P.

XX 07-JUN-2002; 2002US-0386931P.

XX 07-JUN-2002; 2002US-0387078P.

XX 07-JUN-2002; 2002US-0387081P.

XX 07-JUN-2002; 2002US-0387083P.

XX 10-JUN-2002; 2002US-0387429P.

XX 10-JUN-2002; 2002US-0387540P.

XX 10-JUN-2002; 2002US-0387866P.

XX 11-JUN-2002; 2002US-0387606P.

XX 11-JUN-2002; 2002US-0387610P.

XX 11-JUN-2002; 2002US-0387659P.

XX 11-JUN-2002; 2002US-0387668P.

XX 11-JUN-2002; 2002US-0387696P.

XX 11-JUN-2002; 2002US-0387859P.

PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390066P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.
XX (CURA-) CURAGEN CORP.
XX
XX Alsobrook JP, Anderson DM, Baumgartner JC, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Caterton E, Dhanabal M, Edinger SR;
PI Ellerme K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Griesman WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
PI Khrantsov NV, Larochele W, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigar M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernert CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;
XX WPI; 2004-053467/05.
DR P-PSDB; ADH41621.
XX
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.
XX
XX Claim 20; SEQ ID NO 173; 1503pp; English.
PS
XX The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the nucleic acid sequence of the invention.
XX
XX Sequence 2739 BP; 918 A; 550 C; 529 G; 742 T; 0 U; 0 Other;
SQ

Query Match 38.2%; Score 2716.2; DB 12; Length 2739;

Best Local Similarity 99.7%; Pred. No. 0;				Matches 2721; Conservative 0; Mismatches 8; Indels 0; Gaps 0;			
QY	155	CACAGTTGATGTTTCCAAATGTCGTTCTGTGTACTAGTACGATATACCATCTCTTCAA	214				
DB	5	CTCAGGTAGATGTTTCCAAATGTCGTTCTGTGTACTAGTACGATATACCATCTCTTCAA	64				
QY	215	TTTCTACAAACATACACCTCACCTGTTTACTAGAATAGTGACACCAATGTAAACAAACACAG	274				
DB	65	TTTCTACAAACATACACCTCACCTGTTTACTAGAATAGTGACACCAATGTAAACAAACACAG	124				
QY	275	GGCTCTCCAGTCTTCTTAGCCGGGAAAGATCGGATCTCTGTGGGATTTCTTGTCTTGA	334				
DB	125	GGCTCTCCAGTCTTCTTAGCCGGGAAAGATCGGATCTCTGTGGGATTTCTTGTCTTGA	184				
QY	335	ATACACACCTTAATCCAAATGGAAGATATATCTTACATTTGTCAATATTAAGGAAGTTT	394				
DB	185	ATACACACCTTAATCCAAATGGAAGATATATCTTACATTTGTCAATATTAAGGAAGTTT	244				
QY	395	GTCCGTGGATGCAAAACAGTATATACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTC	454				
DB	245	GTCCGTGGATGCAAAACAGTATATACAAAGTCAGATCAAAAGCCAGACAGTCTGGAAGTTC	304				
QY	455	TTCTTTACTAATCTTAATCTCTGGAACAACATATGAAATTTAAGGTTCTGCTGAAACAGTG	514				
DB	305	TTCTTTACTAATCTTAATCTCTGGAACAACATATGAAATTTAAGGTTCTGCTGAAACAGTG	364				
QY	515	CTGGCATTTGGAGTGTGTTAGTGATCAATTTCTTCCAAACCTGCAGAAAGTGTCTCAGGAA	574				
DB	365	CTGGCATTTGGAGTGTGTTAGTGATCAATTTCTTCCAAACCTGCAGAAAGTGTCTCAGGAA	424				
QY	575	RAGTGGTGAATCTCACAGTTGAGGCTACAGCTTACAGCTTACAGCTGATTTGATTT	634				
DB	425	RAGTGGTGAATCTCACAGTTGAGGCTTCAACGCTTACAGCTGATTTGATTTGATTT	484				
QY	635	TACCTCGGCAACCAAAATGCAAAATTTACCAGCTTCAAGATTAGTGTCAAAACATGCCAGAA	694				
DB	485	TACCTCGGCAACCAAAATGCAAAATTTACCAGCTTCAAGATTAGTGTCAAAACATGCCAGAA	544				
QY	695	GTGGGATAGTAGTGAAAGATGTCCTCAATCAGAGTAGAGACATTTTGTAGTGGAAATTCG	754				
DB	545	GTGGGATAGTAGTGAAAGATGTCCTCAATCAGAGTAGAGACATTTTGTAGTGGAAATTCG	604				
QY	755	CAGAATGCAATGAGAAATAGTGAATCTTTTATGAGTACAGCCAGCCCTTCTCCAAACC	814				
DB	605	CAGAATGCAATGAGAAATAGTGAATCTTTTATGAGTACAGCCAGCCCTTCTCCAAACC	664				
QY	815	TTGGTAGAGTTTACACCTCCATCGGTACCAACATTTCAATCAAGCAGTTGTACACAGAAATG	874				
DB	665	TTGGTAGAGTTTACACCTCCATCGGTACCAACATTTCAATCAAGCAGTTGTACACAGAAATG	724				
QY	875	AGATCAGCTCTGTGTGGAAAGAGCTATCAGTTTGTAGTGTGACACATTTGAGACCTTATA	934				
DB	725	AGATCAGCTCTGTGTGGAAAGAGCTATCAGTTTGTAGTGTGACACATTTGAGACCTTATA	784				
QY	935	CAACATATCTTTTGAAGTTTTCAGCTGTACAACTGGAAGCAGGTTTATTTAGTAGTACGA	994				
DB	785	CAACATATCTTTTGAAGTTTTCAGCTGTACAACTGGAAGCAGGTTTATTTAGTAGTACGA	844				
QY	995	TTGTGAGAAACACAGAAATCAGTGTCTGGAAGGACCAACCAAACTGCGTAAACAGGCAACA	1054				
DB	845	TTGTGAGAAACACAGAAATCAGTGTCTGGAAGGACCAACCAAACTGCGTAAACAGGCAACA	904				
QY	1055	TCACAGGAAAGTCTTTTCAATTTTATGGAAGCCCAACCACTATAGTAACAGGGAATTTA	1114				
DB	905	TCACAGGAAAGTCTTTTCAATTTTATGGAAGCCCAACCACTATAGTAACAGGGAATTTA	964				
QY	1115	GTTATAGAGTTGAAATATATATGACCATTCAGGTTCGATTTTGGATTAACAGCAAAAGACC	1174				
DB	965	GTTATAGAGTTGAAATATATATGACCATTCAGGTTCGATTTTGGATTAACAGCAAAAGACC	1024				
QY	1175	TCAAGTTTGCATTTCACTAACCTTAACACCATTTTACAAATTTATGATGTATATTGGCGTGT	1234				

Db 1025 TCAAGTTTGCATTCACCTAACCTAACCAATTTACAAATGATGATGCTATATATGCGGCTG 1084
QY 1235 AAACCAAGTCAGGGAGCTGGGCCCAAGTCAATATTTTCAATTTTCACTCCACCAAGATGTTTC 1294
Db 1085 AAACCAAGTCAGGGAGCTGGGCCCAAGTCAATATTTTCAATTTTCACTCCACCAAGATGTTTC 1144
QY 1295 CAGGGCAGTGTGTTGATTTTCAATCTTGCAGAGTAGAATCCAGCAAGTAAGAAATTTACTT 1354
Db 1145 CAGGGCAGTGTGTTGATTTTCAATCTTGCAGAGTAGAATCCAGCAAGTAAGAAATTTACTT 1204
QY 1355 GGAAGAAACCAACCAACCAATTTGGAATTTAATTAACCAATACCAAGTGAAGTCTAGTTTC 1414
Db 1205 GGAAGAAACCAACCAACCAATTTGGAATTTAATTAACCAATACCAAGTGAAGTCTAGTTTC 1264
QY 1415 CAGAGACAGGAATAATTTTGGAAATACTTTGCTCACTGGAATAATGAGTATATAAATG 1474
Db 1265 CAGAGACAGGAATAATTTTGGAAATACTTTGCTCACTGGAATAATGAGTATATAAATG 1324
QY 1475 ACCCCATGGCTCAGAAATTTGTGAACATAGTAGAGCAATGGTAGAATTTATAGGGTTT 1534
Db 1325 ACCCCATGGCTCAGAAATTTGTGAACATAGTAGAGCAATGGTAGAATTTATAGGGTTT 1384
QY 1535 CAGCAGAGATGTCGTCTGACCTTCACTCTCTCTACATTTATATATAACAGCCATCCAG 1594
Db 1385 CAGCAGAGATGTCGTCTGACCTTCACTCTCTCTACATTTATATATAACAGCCATCCAG 1444
QY 1595 ATAAAAATTTCTCTGCAAGGAATAGAGCTGAAGACCAAGACTTCCAGGTTGTAACACAA 1654
Db 1445 ATAAAAATTTCTCTGCAAGGAATAGAGCTGAAGACCAAGACTTCCAGGTTGTAACACAA 1504
QY 1655 GGAATCAGTATATTTACTGACATTTGCAAGCTGCAAGCTGCTTTATGTTATCAGAGACTTG 1714
Db 1505 GGAATCAGTATATTTACTGACATTTGCAAGCTGCAAGCTGCTTTATGTTATCAGAGACTTG 1564
QY 1715 TACCTTTCTAGTACATGATTTAGTATCTGCTTTTCCATCATGGGAGGAGGACCAC 1774
Db 1565 TACCTTTCTAGTACATGATTTAGTATCTGCTTTTCCATCATGGGAGGAGGACCAC 1624
QY 1775 CAACAGTTCTCAGTGTAGGACACGTCAGCAAGTCGCAAGCTCCATTAATAATTTATAAAT 1834
Db 1625 CAACAGTTCTCAGTGTAGGACACGTCAGCAAGTCGCAAGCTCCATTAATAATTTATAAAT 1684
QY 1835 ATAAAAATTTAGTCTTCACTATTTTGTGTTATTTGGGATCTCCAGAAATATCCCAATG 1894
Db 1685 ATAAAAATTTAGTCTTCACTATTTTGTGTTATTTGGGATCTCCAGAAATATCCCAATG 1744
QY 1895 GAAAAATACCTACTATGCAATTTGCAATGGAATTTGGATACAAACAGAGCATTTCCAGA 1954
Db 1745 GAAAAATACCTACTATGCAATTTGCAATGGAATTTGGATACAAACAGAGCATTTCCAGA 1804
QY 1955 TAACTACCATAGATAACAGCTTTCTCATACAGGGTTTAAAGAAATACACAAATAACAAA 2014
Db 1805 TAACTACCATAGATAACAGCTTTCTCATACAGGGTTTAAAGAAATACACAAATAACAAA 1864
QY 2015 TGAGATGGCAGCTCAACCCACGATGGAAAGTTCTTTGTCGTAAGAAATATGACATCT 2074
Db 1865 TGAGATGGCAGCTCAACCCACGATGGAAAGTTCTTTGTCGTAAGAAATATGACATCT 1924
QY 2075 TTGTGGAATCTCAGAGATGAACCGGAATCATCCTCAAGTGTGGAAGTGAATTTGATG 2134
Db 1925 TTGTGGAATCTCAGAGATGAACCGGAATCATCCTCAAGTGTGGAAGTGAATTTGATG 1984
QY 2135 TTACCCGAGATGAATAAGTTTGAAGTGTGCAACCCGAAAGCCCAATGGGATCATTA 2194
Db 1985 TTACCCGAGATGAATAAGTTTGAAGTGTGCAACCCGAAAGCCCAATGGGATCATTA 2044
QY 2195 TTGCTTATGAAGTGTATATAAATAATAGATATCTTTATATATGAAGAAACACATCAACAA 2254
Db 2045 TTGCTTATGAAGTGTATATAAATAATAGATATCTTTATATATGAAGAAACACATCAACAA 2104
QY 2255 CAGACATATATTAAGGAATTAAGACCTCAGACCTCTATATAACATTTCTGTAAGTCTT 2314
Db 2105 CAGACATATATTAAGGAATTAAGACCTCAGACCTCTATATAACATTTCTGTAAGTCTT 2164

QY 2315 ACAACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGACTTCGGAGA 2374
Db 2165 ACAACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGACTTCGGAGA 2224
QY 2375 CTGTGCTCATATAGTCGACAGAAATATCACTTACAAATATTTCTTCTGGAGAGATTG 2434
Db 2225 CTGTGCTCATATAGTCGACAGAAATATCACTTACAAATATTTCTTCTGGAGAGATTG 2384
QY 2435 AGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTATC 2494
Db 2285 AGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTATC 2344
QY 2495 TCAAGAGAAATGAAGAAATGAGAAAGAACTATAAATCAACCTCTTTAAACCCAAAAACA 2554
Db 2345 TCAAGAGAAATGAAGAAATGAGAAAGAACTATAAATCAACCTCTTTAAACCCAAAAACA 2404
QY 2555 TTAAGGACTGGAAGAAATATACCCATATATCATTTGAGGTGCTGTAGTACACTGAAAG 2614
Db 2405 TTAAGGACTGGAAGAAATATACCCATATATCATTTGAGGTGCTGTAGTACACTGAAAG 2464
QY 2615 GTGAAGGAGTTCGGAGTGTCTCCATAAAGTATATCTACGAGGAGGAAGATGCTCTGATTTCTC 2674
Db 2465 GTGAAGGAGTTCGGAGTGTCTCCATAAAGTATATCTACGAGGAGGAAGATGCTCTGATTTCTC 2524
QY 2675 CCCCTCAAGACTTTCTGTAAACAGTTCTGTGGTGTCAAGGTGAAGTTGTCTAGGCAAC 2734
Db 2525 CCCCTCAAGACTTTCTGTAAACAGTTCTGTGGTGTCAAGGTGAAGTTGTCTAGGCAAC 2584
QY 2735 CACCCCTGAGGCAATGGAATTTATCCCTTTATACACAGTTTATGCTCGAATAGATCAT 2794
Db 2585 CACCCCTGAGGCAATGGAATTTATCCCTTTATACACAGTTTATGCTCGAATAGATCAT 2644
QY 2795 CATTAAAACTAATTAATGTCACTGAAACATCAATGGAGTTATCAGATTTGGATTATAATG 2854
Db 2645 CATTAAAACTAATTAATGTCACTGAAACATCAATGGAGTTATCAGATTTGGATTATAATG 2704
QY 2855 TTGAATACAGTCTTATGTAACAGCTAGC 2883
Db 2705 TTGAATACAGTCTTATGTAACAGCTAGC 2733

RESULT 12

ABT06282

ID ABT06282 standard; cDNA; 2565 BP.

XX AC ABT06282;

XX DT 24-OCT-2002 (first entry)

XX DE Human NOV2b coding sequence.

Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease;
storage disorder; muscle disorder; neurodegenerative disorder; hypotonic;
developmental defect; neuroprotective; antiparkinsonian; hypotensive;
hypertensive; haemostatic; cardiant; antilanginal; dermatological;
immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV;
antiparasitic; antiallergic; antiasthmatic; antirheumatic; antiarthritic;
vulnerary; anorectic; antidiabetic; immunomodulator; antipsoriatic;
nephrotropic; kerolytic; antilucer; cerebroprotective; anticonvulsant;
antinfertility; antilemanic; antidepressant; metabolic; cycostatic;
tranquillizer; analgesic; gene; ss.

OS Homo sapiens.

XX WO200257450-A2.

XX PD 25-JUL-2002.

XX PF 29-NOV-2001; 2001WO-US048922.

XX PR 29-NOV-2000; 2000US-0253834P.

PR 30-NOV-2000; 2000US-0250926P.

XX PA (CURA-) CURAGEN CORP.

XX PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;

XX PI Burgess CE, Casman SU, Catterton E, Dhanabal M, Edinger SR;

PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;

PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;

PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

PI MacLachlan T, Malyankar UM, Mcqueeney K, Mezick AJ, Miller CE;

PI Millet I, Padigaru M, Patturajan M, Peyman J, Qian X, Rascelli L;

PI Rieger DK, Rothenberg MB, Shenoy SG, Shinkets RA, Smithson G;

PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CM, Voss EZ;

PI Wolenc AR, Zhong M, Zhong H;

XX WPI: 2004-053467/05.

DR P-FSDB; ADH41623.

XX

PT New NOVX polypeptides and nucleic acid molecules useful for preventing or

PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,

PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

PT pharmacogenomics.

XX

PS Claim 20; SEQ ID NO 175; 1503pp; English.

XX

CC The invention relates to 566 new isolated human polypeptides and their

CC encoding genes, sequences that are at least 95% identical to these or

CC sequences comprising one or more conservative substitutions in these. The

CC polypeptide, polynucleotide and antibodies against the polypeptides are

CC useful in diagnosing, treating or preventing NOVX-associated disorders,

CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,

CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,

CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.

CC The nucleic acids are further used as hybridization probes, in chromosome

CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The

CC polypeptides are also useful as vaccines. This sequence represents an

CC example of the nucleic acid sequence of the invention.

XX

SQ Sequence 2565 BP; 885 A; 523 C; 447 G; 710 T; 0 U; 0 Other;

Query Match 35.9%; Score 2551; DB 12; Length 2565;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2554; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3262 CCTGAGGGTTGTTGGAACTGCTTACGATCCATTTCTGCTCACTGCTCAATGTA 3321

DB 7 CCTGAGGGTTGTTGGAACTGCTTACGATCCATTTCTGCTCACTGCTCAATGTA 66

QY 3322 AGCTGGGTCCTCCGCTCAACCAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTA 3381

DB 67 AGCTGGGTCCTCCGCTCAACCAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTA 126

QY 3382 CAGCAGACTCCTCGCCATGTGAGACCACTCTGTTTACATATGAGAGAAGCATATATTTT 3441

DB 127 CAGCAGACTCCTCGCCATGTGAGACCACTCTGTTTACATATGAGAGAAGCATATATTTT 186

QY 3442 GATAATCTGGAAAATACACTGATTTATATTAATAAATTTACTCCATCAAGAAAAGGA 3501

DB 187 GATAATCTGGAAAATACACTGATTTATATTAATAAATTTACTCCATCAAGAAAAGGA 246

QY 3502 TTCTCTGATACCTATCTGCTCCAGCTATACATCAAGACTGAAGAAGATGCTCCAGAACT 3561

DB 247 TTCTCTGATACCTATCTGCTCCAGCTATACATCAAGACTGAAGAAGATGCTCCAGAACT 306

QY 3562 TCACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTCTCTTATCATGGGAT 3621

DB 307 TCACCAATAATCAACACTTTTAAAAACCTTTCTCTACCTCAGTCTCTTATCATGGGAT 366

QY 3622 CCCCAGTAAAGCAATGGTGCATTAATAAGTTATGATTTAACTTTAAGGACCAAT 3681

DB 367 CCCCAGTAAAGCAATGGTGCATTAATAAGTTATGATTTAACTTTAAGGACCAAT 426

QY 3682 GAAAATTTCTTCAATCTCTGATAATTAATAATAATTTGAAGAGCTTTTCAACATTT 3741

DB 427 GAAAATTTCTTTCAATTAATCTCTGATAAATACATAAATATTGGAAGAGCTTTTCAACATTT 486

QY 3742 ACATTATATAGCTTTTCTGCGCGCAAGAACTAGAAAAGGACTTTGGTCTCTTCCAGTATT 3801

DB 487 ACATATATATAGCTTTTCTGCGCGCAAGAACTAGAAAAGGACTTTGGTCTCTTCCAGTATT 546

QY 3802 CTTTCTTTTACACAGATGAGTCAGTGGCGTTTAGCACCTCCACAAAATTTGACTTTAATC 3861

DB 547 CTTTCTTTTACACAGATGAGTCAGTGGCGTTTAGCACCTCCACAAAATTTGACTTTAATC 606

QY 3862 AACTGTACTCAGACTTTGTATGCTGGAATGGAGCCCAAGTCTCTTCCAGGTGGTATT 3921

DB 607 AACTGTACTCAGACTTTGTATGCTGGAATGGAGCCCAAGTCTCTTCCAGGTGGTATT 666

QY 3922 GTTAAAGTATATAGTTTAAATTTTCATCAACATGAAACTGCACACTATATATATTAAGAT 3981

DB 667 GTTAAAGTATATAGTTTAAATTTTCATCAACATGAAACTGCACACTATATATTAAGAT 726

QY 3982 ATATCAGGATTTAAATCTGAAGCCAAACTTTGTGGACTGGAAACCAAGTCAGCACCTACTCT 4041

DB 727 ATATCAGGATTTAAATCTGAAGCCAAACTTTGTGGACTGGAAACCAAGTCAGCACCTACTCT 786

QY 4042 ATCCGTGATCTGCTTCCAAAAGTTGGAATGGCAATCAATTTAGTAAATGTAATAAA 4101

DB 787 ATCCGTGATCTGCTTCCAAAAGTTGGAATGGCAATCAATTTAGTAAATGTAATAAA 846

QY 4102 TTCACAAACCCAGGATTCAGTCCAGATCTCGTGCAGAAATATGCAATGCAATGCAACTAGC 4161

DB 847 TTCACAAACCCAGGATTCAGTCCAGATCTCGTGCAGAAATATGCAATGCAATGCAACTAGC 906

QY 4162 TGGCAGTCACTTTTGTAGTGAATGGGATCCACCCAAAAGGCAAAATGGAATTAATAACGAG 4221

DB 907 TGGCAGTCACTTTTGTAGTGAATGGGATCCACCCAAAAGGCAAAATGGAATTAATAACGAG 966

QY 4222 TATATGGTAACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGATCACAATGACT 4281

DB 967 TATATGGTAACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGATCACAATGACT 1026

QY 4282 TTCATAAAGCTTTTGGCAATACCTCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCT 4341

DB 1027 TTCATAAAGCTTTTGGCAATACCTCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCT 1086

QY 4342 GGTGAAGGTGATGAAAGCACATGCTCAGCACACTACCTCGAAACAGTTTCCCAAGTGT 4401

DB 1087 GGTGAAGGTGATGAAAGCACATGCTCAGCACACTACCTCGAAACAGTTTCCCAAGTGT 1146

QY 4402 CCCACAAATATGCTTTTCTGATGTTTCAGTCAACTAGTGCACCAATGCAATGCAATGATA 4461

DB 1147 CCCACAAATATGCTTTTCTGATGTTTCAGTCAACTAGTGCACCAATGCAATGCAATGATA 1206

QY 4462 CTTGACACTATCTTGGCTTACTTTTCAAAAATTTACAAAATTTACCACCTCAACTCGTCTCA 4521

DB 1207 CTTGACACTATCTTGGCTTACTTTTCAAAAATTTACAAAATTTACCACCTCAACTCGTCTCA 1266

QY 4522 AAATGCAAGAATGGGAATCCGAAGAATGTTGTTGAATATCAAAAATTTCAATACCTCTAT 4581

DB 1267 AAATGCAAGAATGGGAATCCGAAGAATGTTGTTGAATATCAAAAATTTCAATACCTCTAT 1326

QY 4582 GAAGCTCACTTAACTGAAGAGACAGTATATGGAATTAAGAAATTTAGATGGTATGATTC 4641

DB 1327 GAAGCTCACTTAACTGAAGAGACAGTATATGGAATTAAGAAATTTAGATGGTATGATTC 1386

QY 4642 GAAGTGGCTGCACAGACCAATGCTGGCTATGGCAATGCTTCAAACTGGATTTCTACAAA 4701

DB 1387 CAAGTGGCTGCACAGACCAATGCTGGCTATGGCAATGCTTCAAACTGGATTTCTACAAA 1446

QY 4702 ACTCTGCTGGCTCCAGATGCTCTCTGAAAATGTTTATGATAGTAGCAACATCACCT 4761

DB 1447 ACTCTGCTGGCTCCAGATGCTCTCTGAAAATGTTTATGATAGTAGCAACATCACCT 1506

QY 4762 TTTAGCATCAGCATAGCTGGAGTGAACCTGCTGCTCATTTATGGAACCAACATGTTATCTG 4821

DB 1507 TTTAGCATCAGCATAGCTGGAGTGAACCTGCTGCTCATTTATGGAACCAACATGTTATCTG 1566

Qy	4822	ATTGATGTCAAATCGGTAGATAATATGATGAATTTTAAATATATCTTTCATCAAGTCCAAATGAA	4881
Db	1567	ATTGATGTCAAATCGGTAGATAATATGATGAATTTTAAATATATCTTTCATCAAGTCCAAATGAA	1626
Qy	4882	GAAATTAATAACCATAGAAAATTAAAGATTTTAAAGAAATTTTCACAAGGTATTTCTGTAGTGCATC	4941
Db	1627	GAAATTAATAACCATAGAAAATTAAAGATTTTAAAGAAATTTTCACAAGGTATTTCTGTAGTGCATC	1686
Qy	4942	ACTGCATTTACTTGGGAACATTTAGTGTGTCATATGTAGAAGGGAAGTCCAAGTGTCTGAAATG	5001
Db	1687	ACTGCATTTACTTGGGAACATTTAGTGTGTCATATGTAGAAGGGAAGTCCAAGTGTCTGAAATG	1746
Qy	5002	ATTGTTACTACTTTTAGAATCAGCCCCAAAGGACCCACTTAACACATGACATTTTCAGAAG	5061
Db	1747	ATTGTTACTACTTTTAGAATCAGCCCCAAAGGACCCACTTAACACATGACATTTTCAGAAG	1806
Qy	5062	ATACCAGATGAAGTTACAAAATTTCAATTAACGTTCTCTCTCTCTCAACCTTAATGGA	5121
Db	1807	ATACCAGATGAAGTTACAAAATTTCAATTAACGTTCTCTCTCTCTCAACCTTAATGGA	1866
Qy	5122	AATATCCAAGTATATCAAGCTCTGGTTTACGGAAGAATGATCTACTGCTGTCCAGATT	5181
Db	1867	AATATCCAAGTATATCAAGCTCTGGTTTACGGAAGAATGATCTACTGCTGTCCAGATT	1926
Qy	5182	CACAACTCTCAGTATTATACAGAAAACCAACACATTTCTGTCATTCGAATGCTTAGAAGACCTA	5241
Db	1927	CACAACTCTCAGTATTATACAGAAAACCAACACATTTCTGTCATTCGAATGCTTAGAAGACCTA	1986
Qy	5242	AAAGGTGGACATACATACAATATCAGTGTGTTTAAGCAGTCAATAGTGTGGTGCAGGTCCA	5301
Db	1987	AAAGGTGGACATACATACAATATCAGTGTGTTTAAGCAGTCAATAGTGTGGTGCAGGTCCA	2046
Qy	5302	AAGTTCGGATGAGAATAACCATGGATATCAAAAGTTCAGCAGCAGCAAAAACCAAAACCA	5361
Db	2047	AAGTTCGGATGAGAATAACCATGGATATCAAAAGTTCAGCAGCAGCAGCAAAAACCAAAACCA	2106
Qy	5362	ACCCCTATTTTATGATGCCACAGGAAAACCTGTTGTGACTTTCAACACAATTTACAATCAGA	5421
Db	2107	ACCCCTATTTTATGATGCCACAGGAAAACCTGTTGTGACTTTCAACACAATTTACAATCAGA	2166
Qy	5422	ATGCCAATATGTTACTACAGTGTATGATCATGGACCAATAAAAAATGTCAAGTGTCTGGC	5481
Db	2167	ATGCCAATATGTTACTACAGTGTATGATCATGGACCAATAAAAAATGTCAAGTGTCTGGC	2226
Qy	5482	ACAGAAACAGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGCATATTTTAAAT	5541
Db	2227	ACAGAAACAGAGCTCAGCATGATGGAATGTAAACAAAGTGTGTATGATGCATATTTTAAAT	2286
Qy	5542	AAAGCAAGGCCATATTTTACAAAATGAAGGCTTTCTTAACCCCTCCATGTACAGAAAGGAAAG	5601
Db	2287	AAAGCAAGGCCATATTTTACAAAATGAAGGCTTTCTTAACCCCTCCATGTACAGAAAGGAAAG	2346
Qy	5602	ACAAAGTTTATGTCGAATGAAGAAATCTACATCATAGGTGTCGATTAATGCATGCATGATT	5661
Db	2347	ACAAAGTTTATGTCGAATGAAGAAATCTACATCATAGGTGTCGATTAATGCATGCATGATT	2406
Qy	5662	CTGTGGCAATGAAGCAAAAATTTTTCGAATGGACCACTGAAACCAAAAAGCAATCTATTATT	5721
Db	2407	CTGTGGCAATGAAGCAAAAATTTTTCGAATGGACCACTGAAACCAAAAAGCAATCTATTATT	2466
Qy	5722	AAATTTTAGAGCTACAAATATTTATGGGACAAATTTACTTGACTCTGATTTATCTGACCCCTGTT	5781
Db	2467	AAATTTTAGAGCTACAAATATTTATGGGCAATTTACTTGACTCTGATTTATCTGACCCCTGTT	2526
Qy	5782	AAGACTTTTAGGGGAAGGACCTTTTCAGAAAGAACCGTAGAG	5820
Db	2527	AAGACTTTTAGGGGAAGGACCTTTTCAGAAAGAACCCCTCGAG	2565

XX	ADH41616;
AC	25-MAR-2004 (first entry)
XX	Novel human nucleic acid NOV15c.
XX	
DE	ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;
XX	anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;
KW	nootropic; antiparkinsonian; antiasthmatic; antifertility;
KW	cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
KW	AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW	Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW	tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX	
OS	Homo sapiens.
XX	
PN	WO2003102159-A2.
XX	
PD	11-DEC-2003.
XX	
PP	04-JUN-2003; 2003WO-US017573.
XX	
PR	04-JUN-2002; 2002US-0385490P.
PR	04-JUN-2002; 2002US-0385615P.
PR	04-JUN-2002; 2002US-0385755P.
PR	05-JUN-2002; 2002US-0386041P.
PR	06-JUN-2002; 2002US-0386355P.
PR	06-JUN-2002; 2002US-0386357P.
PR	06-JUN-2002; 2002US-0386447P.
PR	06-JUN-2002; 2002US-0386459P.
PR	06-JUN-2002; 2002US-0386465P.
PR	06-JUN-2002; 2002US-0386864P.
PR	07-JUN-2002; 2002US-0386701P.
PR	07-JUN-2002; 2002US-0386796P.
PR	07-JUN-2002; 2002US-0386931P.
PR	07-JUN-2002; 2002US-0387081P.
PR	07-JUN-2002; 2002US-0387083P.
PR	10-JUN-2002; 2002US-0387429P.
PR	10-JUN-2002; 2002US-0387540P.
PR	10-JUN-2002; 2002US-0387866P.
PR	11-JUN-2002; 2002US-0387606P.
PR	11-JUN-2002; 2002US-0387610P.
PR	11-JUN-2002; 2002US-0387659P.
PR	11-JUN-2002; 2002US-0387668P.
PR	11-JUN-2002; 2002US-0387696P.
PR	11-JUN-2002; 2002US-0387859P.
PR	12-JUN-2002; 2002US-0387934P.
PR	12-JUN-2002; 2002US-0387960P.
PR	12-JUN-2002; 2002US-0388022P.
PR	12-JUN-2002; 2002US-0388096P.
PR	12-JUN-2002; 2002US-0388432P.
PR	12-JUN-2002; 2002US-0388479P.
PR	13-JUN-2002; 2002US-0389123P.
PR	14-JUN-2002; 2002US-0389120P.
PR	14-JUN-2002; 2002US-0389146P.
PR	17-JUN-2002; 2002US-0389742P.
PR	18-JUN-2002; 2002US-0389604P.
PR	18-JUN-2002; 2002US-0389884P.
PR	19-JUN-2002; 2002US-0390066P.
PR	19-JUN-2002; 2002US-0390144P.
PR	19-JUN-2002; 2002US-0390209P.
PR	25-JUN-2002; 2002US-0391726P.
PR	06-AUG-2002; 2002US-0401628P.
PR	09-AUG-2002; 2002US-0402268P.
PR	12-AUG-2002; 2002US-0402822P.
PR	13-AUG-2002; 2002US-0403458P.
PR	15-AUG-2002; 2002US-0403617P.
PR	15-AUG-2002; 2002US-0403732P.
PR	26-AUG-2002; 2002US-0406182P.
PR	12-SEP-2002; 2002US-0410085P.
PR	13-SEP-2002; 2002US-0410505P.


```
QY 2458 AGTCCCAATGGAATCATATAAATAATATACAAATTTATCTCAAGAGAAATGTAATGGAATGAG 2517
Db 1447 AGTCCCAATGGAATCATATAAATAATATACAAATTTATCTCAAGAGAAATGTAATGGAATGAG 1506
QY 2518 GAAAGAACTATAATAACAACTCTTTAAACCCAAACATTAAGTACTTGAAGAATAATATACC 2577
Db 1507 GAAAGAACTATAATAACAACTCTTTAAACCCAAACATTAAGTACTTGAAGAATAATATACC 1566
QY 2578 CAATATATCATTCAGGTGCTCTGTAGTACACTCAAAAGGTGAAGGAGTTCGGAGTCTCC 2637
Db 1567 CAATATATCATTCAGGTGCTCTGTAGTACACTCAAAAGGTGAAGGAGTTCGGAGTCTCC 1626
QY 2638 ATAACTATATCTGACGAGGAGAGATGCTCTGTATCTCCCTCAAGACTTCTCTGTAA 2697
Db 1627 ATAACTATATCTGACGAGGAGAGATGCTCTGTATCTCCCTCAAGACTTCTCTGTAA 1686
QY 2698 CAGTGTCTGTCTCAGGTGAAGTGTCTATGCAACCAACCCCTGAGGCAAAATGGAAT 2757
Db 1687 CAGTGTCTGTCTCAGGTGAAGTGTCTATGCAACCAACCCCTGAGGCAAAATGGAAT 1746
QY 2758 ATCTTTTATPACACAGTTTATGTCTGGAATAGATCATATTAATAAACTATTAATGTCAC 2817
Db 1747 ATCTTTTATPACACAGTTTATGTCTGGAATAGATCATATTAATAAACTATTAATGTCAC 1806
QY 2818 GAAACATCATTTGAGTTATCAGATTTTGGATTAATGTTGAATACAGTGTATGTAACA 2877
Db 1807 GAAACATCATTTGAGTTATCAGATTTTGGATTAATGTTGAATACAGTGTATGTAACA 1866
QY 2878 GCTAGCACCAAGATTTCTGTATGGAAGGAAACAGGAAGCAATATCATTAGCTTCCAAACCA 2937
Db 1867 GCTAGCACCAAGATTTCTGTATGGAAGGAAACAGGAAGCAATATCATTAGCTTCCAAACCA 1926
QY 2938 GAGGAGCACCAAGCGATCTCCCAAGATGTTTATATGCAAACTCAGTCTTTCATCA 2997
Db 1927 GAGGAGCACCAAGCGATCTCCCAAGATGTTTATATGCAAACTCAGTCTTTCATCA 1986
QY 2998 ATAAATCTTTCTGGAACCTCTTCAAACTTAATGGAATATACAAATATCTCTGTT 3057
Db 1987 ATAAATCTTTCTGGAACCTCTTCAAACTTAATGGAATATACAAATATCTCTGTT 2046
QY 3058 TATTACAGAAATCTTCAGTACTTTTATGCAAACTTATGCAAACTTATGCAAACTTATGCAAACT 3117
Db 2047 TATTACAGAAATCTTCAGTACTTTTATGCAAACTTATGCAAACTTATGCAAACTTATGCAAACT 2106
QY 3118 GACTTTGACAAATGACTGTATCCCAATATATAGATAAACTGCAAACTTATGCAAACTTATGCAAACT 3177
Db 2107 GACTTTGACAAATGACTGTATCCCAATATATAGATAAACTGCAAACTTATGCAAACTTATGCAAACT 2166
QY 3178 ACATTTTGGTTAAACAGCAAGTACTTCAAGTTGGAATGGAATTAAGGAGTGA 3230
Db 2167 ACATTTTGGTTAAACAGCAAGTACTTCAAGTTGGAATGGAATTAAGGAGTGA 2219
```

RESULT 15

AAQ73786

ID AAQ73786 standard; cDNA; 2309 BP.

XX AC AAQ73786;

XX 25-MAR-2003 (revised)

DT 26-JUN-1995 (first entry)

XX Partial PCR fragment of PTP-S31D.

DE Protein tyrosine phosphatase; cancer; diabetes; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 65..1035

FT /tag= a

FT misc_feature 1766

/*tag= b
/notes= "unknown base"

W09421800-A2.

29-SEP-1994.

23-MAR-1994; 94WO-BP000909.

23-MAR-1993; 93US-00036210.

(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Moller NPH, Moller KB, Ullrich A;

WPI; 1994-317020/39.

P-PSDB; AAR60877.

New protein tyrosine phosphatase PTP-S31 - is used to develop prods. for

treating or preventing disease associated with abnormal PTP-S31, e.g.

cancer or diabetes.

Claim 6; Fig 6; 116pp; English.

The sequence is that of a partial fragment of a new protein tyrosine

phosphatase PTP-S31D that was isolated from RNA from human skeletal

muscle. Such DNA can be used to treat cancer and diabetes. See also

AAQ73782-8. (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 2309 BP; 770 A; 384 C; 435 G; 719 T; 0 U; 1 Other;

Query Match 15.1%; Score 1071.6; DB 2; Length 2309;

Best Local Similarity 99.4%; Pred. No. 1.4e-245;

Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 6015 AACGGGCAATTAAGCAAGAAATCTCTCTGCAATGTTGAGAGCTTTGCAACAACAA 6074

Db 70 AATGAGGCAATTAAGCAAGAAATCTCTCTGCAATGTTGAGAGCTTTGCAACAACAA 129

QY 6075 CAACCTAAAGTTTCAAGAAAGAAATTTTGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC 6134

Db 130 CAACCTAAAGTTTCAAGAAAGAAATTTTGGAAATTTACCAAAATTTCTTCAGGATCTTTCTTC 189

QY 6135 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCGTTTCCAAAACATAAAACATA 6194

Db 190 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCGTTTCCAAAACATAAAACATA 248

QY 6195 TAATTAATTAACAGAGTAAGCTGATAGCTGACGCTAGTGTCCAGGTTCCGATTTATAT 6254

Db 249 --ATAATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTCCAGGTTCCGATTTATAT 306

QY 6255 TAATGCGAGCTATATTTCTGGTTTATTTATGTCAAATGAATTTATTTGCTACTCAAGGTC 6314

Db 307 TAATGCGAGCTATATTTCTGGTTTATTTATGTCAAATGAATTTATTTGCTACTCAAGGTC 366

QY 6315 ACTACAGAGCAAGTTGGAGATTTTGGAGAAATGTTGGGAAACCCAGGCGCAAAACATTT 6374

Db 367 ACTACAGAGCAAGTTGGAGATTTTGGAGAAATGTTGGGAAACCCAGGCGCAAAACATTT 426

QY 6375 AGTAATGCTAACACAGCTGTTTGAAGAGGAGGATCAGATCCCATCAGTATTTGGCCAGA 6434

Db 427 AGTAATGCTAACACAGCTGTTTGAAGAGGAGGATCAGATCCCATCAGTATTTGGCCAGA 486

QY 6435 GGACAAACAGCCAGTTTACTGCTTTGGAGATATAGTATTACAAAGCTAATAGGAGATGT 6494

Db 487 GGACAAACAGCCAGTTTACTGCTTTGGAGATATAGTATTACAAAGCTAATAGGAGATGT 546

QY 6495 TCAAAATAGATTGGACTATCAGGGATCTGAAAATGGAAGGCAATGGGATTTGATGACTGT 6554

Db 547 TCAAAATAGATTGGACTATCAGGGATCTGAAAATGGAAGGCAATGGGATTTGATGACTGT 606

QY 6555 TCGACAGTGTAACTTTACTGCTGCGCCAGAGATGGGTTCTTGAGAACAGGCCCTCTCT 6614

Db 607 TCGACAGTGTAACTTTACTGCTGGCCAGAGCATGGGTTCTGTAGAAACAGCGCCCTCT 666
Qy 6615 AATTCACTTTTGAAGTTGGTTGAGCAAGCAGGGCACAATGACACACACCTATGATTGT 6674
Db 667 AATTCACTTTTGAAGTTGGTTGAGCAAGCAGGGCACAATGACACACACCTATGATTGT 726
Qy 6675 TCACTGCAGTGTGGAGTTGGAGAACTCGAGTGTATTTATTTGCTCTGGACCAATTTAACACA 6734
Db 727 TCACTGCAGTGTGGAGTTGGAGAACTCGAGTGTATTTATTTGCTCTGGACCAATTTAACACA 786
Qy 6735 ACATATAAATGACCAATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGAGAAGTGA 6794
Db 787 ACATATAAATGACCAATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGAGAAGTGA 846
Qy 6795 AGAATGTGATGTGCGAGAACTGCGCAGATATATCTTTTACACCACTGCAATCTGGA 6854
Db 847 AAGAATGTGATGTGCGAGAACTGCGCAGATATATCTTTTACACCACTGCAATCTGGA 906
Qy 6855 TCTCTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCACTTCA 6914
Db 907 TCTCTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCACTTCA 966
Qy 6915 GAAGATGGACTCTTTTGGACGCCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCAC 6974
Db 967 GAAGATGGACTCTTTTGGACGCCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAAACCAC 1026
Qy 6975 TATGTAATAATTCAGACCAAGGATACAAATGGAAGAGATTTTAAATCCAGGGGCCAA 7034
Db 1027 TATGTAATAATTCAGACCAAGGATACAAATGGAAGAGATTTTAAATCCAGGGGCCAA 1086
Qy 7035 AGTTACCCCTCATCTTCCGAATTCGAATGTCAACCTTAAAGAAATATCTATGCTTCT 7094
Db 1087 AGTTACCCCTCATCTTCCGAATTCGAATGTCAACCTTAAAGAAATATCTATGCTTCT 1146
Qy 7095 CTCACCTGTGCTTT 7108
Db 1147 CTCACCTGTGCTTT 1160

Search completed: July 1, 2005, 22:46:26
Job time : 5236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 21:03:34 ; Search time 1608 Seconds
(without alignments)
7232.994 Million cell updates/sec

Title: US-10-673-885-1
Perfect score: 7108
Sequence: 1 taattgtacttgccagaa.....gcttctcactgtgcttt 7108

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7108	100.0	7108	4	US-09-822-871-1
2	2618.6	36.8	3973	1	US-08-036-210-21
3	2618.6	36.8	3973	2	US-08-449-609-21
4	2618.6	36.8	3973	4	US-09-361-096A-21
5	1118.8	15.7	2692	1	US-08-036-210-14
6	1118.8	15.7	2692	2	US-08-449-609-14
7	1118.8	15.7	2692	4	US-09-361-096A-14
8	1071.6	15.1	2309	1	US-08-036-210-10
9	1071.6	15.1	2309	2	US-08-449-609-10
10	1071.6	15.1	2309	4	US-09-361-096A-10
11	789.6	11.1	2173	1	US-08-036-210-6
12	789.6	11.1	2173	2	US-08-449-609-6
13	789.6	11.1	2173	4	US-09-361-096A-6
14	532	7.5	532	1	US-08-036-210-8
15	532	7.5	532	2	US-08-449-609-8
16	532	7.5	532	4	US-09-361-096A-8
17	319.4	4.5	321	1	US-08-036-210-3
18	319.4	4.5	321	2	US-08-449-609-3
19	319.4	4.5	321	4	US-09-361-096A-3
20	315	4.4	254366	4	US-09-822-871-3
21	195.2	2.7	5679	1	US-08-201-697-1
22	192.2	2.7	4064	4	US-09-949-016-2962
23	192.2	2.7	4064	4	US-09-949-016-2963
24	192.2	2.7	4064	4	US-09-949-016-2964
25	192.2	2.7	4064	4	US-09-949-016-2965
26	192.2	2.7	4064	4	US-09-949-016-2966
27	192.2	2.7	4064	4	US-09-949-016-2967

28	192.2	2.7	4752	1	US-08-201-697-3	Sequence 3, Appli
29	192.2	2.7	4815	1	US-08-201-697-5	Sequence 5, Appli
30	192.2	2.7	5069	4	US-09-949-016-2180	Sequence 2180, Ap
31	192.2	2.7	5069	4	US-09-949-016-2181	Sequence 2181, Ap
32	192.2	2.7	5069	4	US-09-949-016-2182	Sequence 2182, Ap
33	192.2	2.7	5069	4	US-09-949-016-2183	Sequence 2183, Ap
34	192.2	2.7	5069	4	US-09-949-016-2184	Sequence 2184, Ap
35	192.2	2.7	5153	4	US-09-949-016-2957	Sequence 2957, Ap
36	192.2	2.7	5153	4	US-09-949-016-2958	Sequence 2958, Ap
37	192.2	2.7	5153	4	US-09-949-016-2959	Sequence 2959, Ap
38	192.2	2.7	5153	4	US-09-949-016-2960	Sequence 2960, Ap
39	192.2	2.7	5153	4	US-09-949-016-2961	Sequence 2961, Ap
40	192.2	2.7	6075	4	US-09-023-655-1366	Sequence 1366, Ap
41	192.2	2.7	6075	4	US-09-949-016-404	Sequence 404, App
42	191.6	2.7	5763	4	US-09-949-016-2178	Sequence 2178, Ap
43	187.4	2.6	4799	1	US-08-201-697-6	Sequence 6, Appli
44	171.8	2.4	4756	4	US-09-023-655-1461	Sequence 1461, Ap
45	171.8	2.4	5117	3	US-08-854-585-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-822-871-1
; Sequence 1, Application US/09822871
; Patent No. 6723547
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: CL001219
; CURRENT APPLICATION NUMBER: US/09/822,871
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 7108
; TYPE: DNA
; ORGANISM: Human
US-09-822-871-1

Query Match	100.0%	Score 7108;	DB 4;	Length 7108;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 7108;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TAATTGTGTACTGCGAGAGGATCTGCTTTAAATCATTATGAGGCAACATTTCCT	60	
Db	1	TAATTGTGTACTGCGAGAGGATCTGCTTTAAATCATTATGAGGCAACATTTCCT	60	
QY	61	CTAGAGCCATCAATGTGTTCTACTGGCTGAAATAATGTAATAAGATGATTTCTTATC	120	
Db	61	CTAGAGCCATCAATGTGTTCTACTGGCTGAAATAATGTAATAAGATGATTTCTTATC	120	
QY	121	ATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGATTTTCCAATGTCGTT	180	
Db	121	ATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTTGATTTTCCAATGTCGTT	180	
QY	181	CTGTGTACTAGTAGATATAACCATCTTCTTCAATTTCTAACACATACACCTCCTGTT	240	
Db	181	CTGTGTACTAGTAGATATAACCATCTTCTTCAATTTCTAACACATACACCTCCTGTT	240	
QY	241	ACTAGATAGTCAGACCAATGTAACAAACAGGCTTCTTCTAGCCGGGAA	300	
Db	241	ACTAGATAGTCAGACCAATGTAACAAACAGGCTTCTTCTAGCCGGGAA	300	
QY	301	AGAGTCGGATCTGCTGGGATTTCTTCTGTCTTGAATACACACCTAATCAATGAAGG	360	
Db	301	AGAGTCGGATCTGCTGGGATTTCTTCTGTCTTGAATACACACCTAATCAATGAAGG	360	
QY	361	ATTATATCTTACATTTGTCAATATATAGGAAGTTTGTCCGTGGATGCAACAGTATATACA	420	
Db	361	ATTATATCTTACATTTGTCAATATATAGGAAGTTTGTCCGTGGATGCAACAGTATATACA	420	

Db 361 ATTATATCTTACATTTGTCAAATATAAGGAAGTTTGTCCGTGGATGCAAAACAGTATATACA 420
Qy 421 CAGAGTCAGATCAAGCCAGACAGCTCTGGAAGTTCTTCTTACTTAATCTTAACTCTTGGAACA 480
Db 421 CAAAGTCAGATCAAAAGCCAGACAGCTCTGGAAGTTCTTCTTACTTAATCTTAACTCTTGGAACA 480
Qy 481 ACATATGAAATTAAGGTTGCTGCTGAAACACAGTCTGGCATTTGGAGTGTTTAGTGATCCA 540
Db 481 ACATATGAAATTAAGGTTGCTGCTGAAACACAGTCTGGCATTTGGAGTGTTTAGTGATCCA 540
Qy 541 TTTCTCTTCCAACTGCGAAGTGTCTCCAGGAAAGTGGTGAATCTCAGFTGAGGCC 600
Db 541 TTTCTCTTCCAACTGCGAAGTGTCTCCAGGAAAGTGGTGAATCTCAGFTGAGGCC 600
Qy 601 TACAAAGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCAACCAATGGCAAAATTT 560
Db 601 TACAAAGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCAACCAATGGCAAAATTT 560
Qy 661 ACCAGCTTCAAGATTAAGTGTCAAACATGCCAGAAAGTGGATAGTAGTGAAGATGTCTCA 720
Db 661 ACCAGCTTCAAGATTAAGTGTCAAACATGCCAGAAAGTGGATAGTAGTGAAGATGTCTCA 720
Qy 721 ATCAGAGTAGAGCAATTTTGAATCGGAAATTCAGAAATGCCAAGATGAGATAGTGAATCT 780
Db 721 ATCAGAGTAGAGCAATTTTGAATCGGAAATTCAGAAATGCCAAGATGAGATAGTGAATCT 780
Qy 781 TTTTATGAGTAGACGCGAGCCCTTCTCCAACTTGGTAGAGTTACACTCCATCGCCT 840
Db 781 TTTTATGAGTAGACGCGAGCCCTTCTCCAACTTGGTAGAGTTACACTCCATCGCCT 840
Qy 841 ACCACACATTTCAAGCAGCTGTGACACAGAAATGAGATCAGCTCTGTGTGGAAGAGCCT 900
Db 841 ACCACACATTTCAAGCAGCTGTGACACAGAAATGAGATCAGCTCTGTGTGGAAGAGCCT 900
Qy 901 ATCAGTTTTGTAGTAGACACATTTGAGACCTTATACAACATATCTTTTGAAGTTTCAGCT 960
Db 901 ATCAGTTTTGTAGTAGACACATTTGAGACCTTATACAACATATCTTTTGAAGTTTCAGCT 960
Qy 961 GCTTACAACTGMAAGCAGTTATATTGATAGTAGACATTTGTGAGAACACAGAACTCAGTGCCT 1020
Db 961 GCTTACAACTGMAAGCAGTTATATTGATAGTAGACATTTGTGAGAACACAGAACTCAGTGCCT 1020
Qy 1021 GAAGGACCCACCAAACTGCGTAAACAGGCAACATCAAGGAAAGTCTTTTCAATTTTA 1080
Db 1021 GAAGGACCCACCAAACTGCGTAAACAGGCAACATCAAGGAAAGTCTTTTCAATTTTA 1080
Qy 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTAATGAGCCA 1140
Db 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTAATGAGCCA 1140
Qy 1141 TCAGGTGCGATTTTGGATTAACAGCAAAAAGACCTCAAGTTTGCATTTCACTAACCTTAAACA 1200
Db 1141 TCAGGTGCGATTTTGGATTAACAGCAAAAAGACCTCAAGTTTGCATTTCACTAACCTTAAACA 1200
Qy 1201 CCATTTACAAATGATGATGCTATATTGCGGCTGAAACACAGTGCAGGGACTGCGGCCAAG 1260
Db 1201 CCATTTACAAATGATGATGCTATATTGCGGCTGAAACACAGTGCAGGGACTGCGGCCAAG 1260
Qy 1261 TCAAAATTTTCAATTTCACTCCACAGATTTTCCAGGGGCGAGTGTTCGATTTTCAACTTT 1320
Db 1261 TCAAAATTTTCAATTTCACTCCACAGATTTTCCAGGGGCGAGTGTTCGATTTTCAACTTT 1320
Qy 1321 GCAGAGGTAGATTCACGCAAGTAAGAAATTTACTTTGGAAGAAACCCAGACAAACCAATGGA 1380
Db 1321 GCAGAGGTAGATTCACGCAAGTAAGAAATTTACTTTGGAAGAAACCCAGACAAACCAATGGA 1380
Qy 1381 ATTTATTAACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATATTTTGGAAAT 1440
Db 1381 ATTTATTAACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATATTTTGGAAAT 1440
Qy 1441 ACTTTGCTCACTGGAATTAATGAGTATATAAATGACCCCACTGCTCCAGAAATTTGTGAAC 1500
Db 1441 ACTTTGCTCACTGGAATTAATGAGTATATAAATGACCCCACTGCTCCAGAAATTTGTGAAC 1500

Qy 1501 ATAGTAGAGCCCAATGTTAGGATTTATATGAGGTTTCAGCAGAGATGTCGTCTGACCTTCAC 1560
Db 1501 ATAGTAGAGCCCAATGTTAGGATTTATATGAGGTTTCAGCAGAGATGTCGTCTGACCTTCAC 1560
Qy 1561 TCATTTGCTTACATTTTATATATAACAGCCATCCAGATAAATACTTCTGCAAGGAATAGA 1620
Db 1561 TCATTTGCTTACATTTTATATATAACAGCCATCCAGATAAATACTTCTGCAAGGAATAGA 1620
Qy 1621 GCTGAAGACCCAGACTTTCACAGTTGTAACTACAGGAATCAGTATATTACTGACATTGCA 1680
Db 1621 GCTGAAGACCCAGACTTTCACAGTTGTAACTACAGGAATCAGTATATTACTGACATTGCA 1680
Qy 1681 GCTGAACAGCTGCTTTATGTTTATCAGGAGACTTGTACCTTTTCACTGAGCACAATGATCT 1740
Db 1681 GCTGAACAGCTGCTTTATGTTTATCAGGAGACTTGTACCTTTTCACTGAGCACAATGATCT 1740
Qy 1741 GTATCTGCTTTTCCATCATGCGGAGAGGACCAACCAAGCTTCTCAGTGTTTAGGACACGT 1800
Db 1741 GTATCTGCTTTTCCATCATGCGGAGAGGACCAACCAAGCTTCTCAGTGTTTAGGACACGT 1800
Qy 1801 CAGCAAGTGCAGAGCTCCATTTAAATTTATAAACTATATAAAATATTAGTCTTCTCATCTAT 1860
Db 1801 CAGCAAGTGCAGAGCTCCATTTAAATTTATAAACTATATAAAATATTAGTCTTCTCATCTAT 1860
Qy 1861 TTTGTTTATATTGGGATCCTCCAGATATCCCAATGAAAAATAAATCACTATACGATTTAT 1920
Db 1861 TTTGTTTATATTGGGATCCTCCAGATATCCCAATGAAAAATAAATCACTATACGATTTAT 1920
Qy 1921 GCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAAATACCATAGATAAAGCTTTCTC 1980
Db 1921 GCAATGGAATTTGGATACAAACAGAGCATTTCCAGATAAATACCATAGATAAAGCTTTCTC 1980
Qy 1981 ATAAACAGGTTTAAAGAAATACAAAATACAAAATGAGAGTGCGAGCCTCAACCCACGAT 2040
Db 1981 ATAAACAGGTTTAAAGAAATACAAAATGAGAGTGCGAGCCTCAACCCACGAT 2040
Qy 2041 GAGGAAAGTTCTTGTCTGGAAGAAATGACATCTTTGTGAGAACTTTCAGAGATGAACG 2100
Db 2041 GAGGAAAGTTCTTGTCTGGAAGAAATGACATCTTTGTGAGAACTTTCAGAGATGAACG 2100
Qy 2101 GAATCATCACTTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAATAAGTTGAAG 2160
Db 2101 GAATCATCACTTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAATAAGTTGAAG 2160
Qy 2161 TGGTCAACCCGAAAGCCCAATGGGATCATTTGCTTATGAGAGTGTATATAAAT 2220
Db 2161 TGGTCAACCCGAAAGCCCAATGGGATCATTTGCTTATGAGAGTGTATATAAAT 2220
Qy 2221 ATAGATCTTTTATATATGAAGAAACATCAACAAACAGACATATAATTAAGGAACCTTAAGA 2280
Db 2221 ATAGATCTTTTATATATGAAGAAACATCAACAAACAGACATATAATTAAGGAACCTTAAGA 2280
Qy 2281 CCTCACACCTCTATAACATTTCTGTAAAGTCTTACACAGATTTGGTCAATGCAATCAG 2340
Db 2281 CCTCACACCTCTATAACATTTCTGTAAAGTCTTACACAGATTTGGTCAATGCAATCAG 2340
Qy 2341 GTATCTTCTTACTCTCTGTAAAGGACTTCGGAGACTGTGCGCTGATGTCACAGAAAT 2400
Db 2341 GTATCTTCTTACTCTCTGTAAAGGACTTCGGAGACTGTGCGCTGATGTCACAGAAAT 2400
Qy 2401 ATCAGCTTCAAAAATATTTCTTCGAGAGATTTGAGCTATCATTTCTTCCCTCCCAAGTAGT 2460
Db 2401 ATCAGCTTCAAAAATATTTCTTCGAGAGATTTGAGCTATCATTTCTTCCCTCCCAAGTAGT 2460
Qy 2461 CCCAATGGAATCATTAATAAATAATACAAATTTATCTCAAGAGAAAGTAATGGAATGAGGAA 2520
Db 2461 CCCAATGGAATCATTAATAAATAATACAAATTTATCTCAAGAGAAAGTAATGGAATGAGGAA 2520
Qy 2521 AGAAGCTATAAATCAAACTCTTTTAAACCCAAACATTAAGTACTGAAAGAAATATATACCCA 2580
Db 2521 AGAAGCTATAAATCAAACTCTTTTAAACCCAAACATTAAGTACTGAAAGAAATATATACCCA 2580

QY 2581 TATATCAATTCAGGTGTCCTGCTAGTACACTGAAAGGTGAGAGCTTCGGAGTGCTCCCAT 2640
Db 2581 TATATCAATTCAGGTGTCCTGCTAGTACACTGAAAGGTGAGAGCTTCGGAGTGCTCCCAT 2640
QY 2641 AGTATACAGCGGAGGAAGATGCTCCTGATTCCTCCCTCAAGACTTCTCTGTAAACAG 2700
Db 2641 AGTATACAGCGGAGGAAGATGCTCCTGATTCCTCCCTCAAGACTTCTCTGTAAACAG 2700
QY 2701 TTGTCGGTGTCAAGGTGAAGTGTGATGCGCAACCAACCCCTGGAGGCAAAATGGAATATC 2760
Db 2701 TTGTCGGTGTCAAGGTGAAGTGTGATGCGCAACCAACCCCTGGAGGCAAAATGGAATATC 2760
QY 2761 CTTTATTACACAGTTTATGTCGGAATAGATCATTAATAAATACTAATGTCACGTGA 2820
Db 2761 CTTTATTACACAGTTTATGTCGGAATAGATCATTAATAAATACTAATGTCACGTGA 2820
QY 2821 ACATCAATGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGTCTATGTAACAGCT 2880
Db 2821 ACATCAATGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGTCTATGTAACAGCT 2880
QY 2881 AGCACAGATTTGGTGAATGGGAAAAAGGAAGCAATATCATATTAGCTTTCAACACAGAG 2940
Db 2881 AGCACAGATTTGGTGAATGGGAAAAAGGAAGCAATATCATATTAGCTTTCAACACAGAG 2940
QY 2941 GGAAGCAAGAGCATCTCCCAAGATGTTTATTATGCAAACTCAGTTCCTTCATCAATA 3000
Db 2941 GGAAGCAAGAGCATCTCCCAAGATGTTTATTATGCAAACTCAGTTCCTTCATCAATA 3000
QY 3001 ATTCTTTCTGACACCTCTCTCAAACTTAATGGGATATATAAATATTACTCTGTTTAT 3060
Db 3001 ATTCTTTCTGACACCTCTCTCAAACTTAATGGGATATATAAATATTACTCTGTTTAT 3060
QY 3061 TACAGAAATACTTCAGGTACTTTTATGCGAAATTTTACACTCCATGAATCAACCAATGAC 3120
Db 3061 TACAGAAATACTTCAGGTACTTTTATGCGAAATTTTACACTCCATGAATCAACCAATGAC 3120
QY 3121 TTGACAAATAGCTGTATCCCAATATAGATAAATGACAAATATTCAGTACTATACA 3180
Db 3121 TTGACAAATAGCTGTATCCCAATATAGATAAATGACAAATATTCAGTACTATACA 3180
QY 3181 TTTTGGTTAAACAGCAAGTACTTCAGTTGGAATGGGAATAAAGCAGTACATTTGAA 3240
Db 3181 TTTTGGTTAAACAGCAAGTACTTCAGTTGGAATGGGAATAAAGCAGTACATTTGAA 3240
QY 3241 GTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAAACCTGACTTACGAATCCAAT 3300
Db 3241 GTATACACAGATCAAGACATACCTGAAAGGTTTGTGGAAACCTGACTTACGAATCCAAT 3300
QY 3301 TCGTCAACTGCAATAAATGTAAGCTGGTCCCAACCGGCTCAACCAACGGTCTAGTCTTC 3360
Db 3301 TCGTCAACTGCAATAAATGTAAGCTGGTCCCAACCGGCTCAACCAACGGTCTAGTCTTC 3360
QY 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCGGCATGTGAGACCACTCTGTTTACA 3420
Db 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCGGCATGTGAGACCACTCTGTTTACA 3420
QY 3421 TATGAGAGAGCATATATTTGATAATCTGGAAAAATACATGATTATATATAAATAAT 3480
Db 3421 TATGAGAGAGCATATATTTGATAATCTGGAAAAATACATGATTATATATAAATAAT 3480
QY 3481 ACTCCATCAACGAAAGGGAATTCCTGTAATCCTATATGCCCCAGCTATACATCAAGACT 3540
Db 3481 ACTCCATCAACGAAAGGGAATTCCTGTAATCCTATATGCCCCAGCTATACATCAAGACT 3540
QY 3541 GAAGAAGATGTCACGAAACTTCAACCAATAATCAACACTTTTAAACACCTTCTCTACC 3600
Db 3541 GAAGAAGATGTCACGAAACTTCAACCAATAATCAACACTTTTAAACACCTTCTCTACC 3600
QY 3601 TCAGTTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGCAATAATAAGTTATGAT 3660
Db 3601 TCAGTTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGCAATAATAAGTTATGAT 3660
QY 3661 TTAACCTTTACAGGACCAAAATGAAATTTATTTCTTCAATTAATTCATTAATAATAATA 3720

Db 3661 TTAACCTTTACAGGACCAAAATGAAATTTATTTCTTCAATTAATTCATTAATAATAATA 3720
QY 3721 TTGGAAGAGCTTTCAACCATTTATATATAGCTTTTTTGTGCGCAAGAACTAGAAAA 3780
Db 3721 TTGGAAGAGCTTTCAACCATTTATATATAGCTTTTTTGTGCGCAAGAACTAGAAAA 3780
QY 3781 GGAAGAGCTTTCAACCATTTATATATAGCTTTTTTGTGCGCAAGAACTAGAAAA 3840
Db 3781 GGAAGAGCTTTCAACCATTTATATATAGCTTTTTTGTGCGCAAGAACTAGAAAA 3840
QY 3841 GGAAGAGCTTTCAACCATTTATATATAGCTTTTTTGTGCGCAAGAACTAGAAAA 3900
Db 3841 GGAAGAGCTTTCAACCATTTATATATAGCTTTTTTGTGCGCAAGAACTAGAAAA 3900
QY 3901 AGTCTCTTTCCAGAGTGTATTTGTAAGATATATAGTTTTTAAATTCATGAATGAACT 3960
Db 3901 AGTCTCTTTCCAGAGTGTATTTGTAAGATATATAGTTTTTAAATTCATGAATGAACT 3960
QY 3961 GACACTATATATATAAGAAATATATCAGGATTTAAACTGAAAGCCAAACTTTGTTGACTG 4020
Db 3961 GACACTATATATATAAGAAATATATCAGGATTTAAACTGAAAGCCAAACTTTGTTGACTG 4020
QY 4021 GAAACAGTCAAGCACTCTATCCGCTGATCTATCCGCTGATCTGCGTTCACCAAGTTGGAATGCAAT 4080
Db 4021 GAAACAGTCAAGCACTCTATCCGCTGATCTATCCGCTGATCTGCGTTCACCAAGTTGGAATGCAAT 4080
QY 4081 CAATTTAGTAATGTAAGAAATTTCAACCCCAAGAAATCAGTTCCAGATGCTGTCAGAAAT 4140
Db 4081 CAATTTAGTAATGTAAGAAATTTCAACCCCAAGAAATCAGTTCCAGATGCTGTCAGAAAT 4140
QY 4141 ATGAGTGTGCAAGCACTAGCTGCGAGTCAAGTTTGTAGTGAATGGATCCACCAAGAAAT 4200
Db 4141 ATGAGTGTGCAAGCACTAGCTGCGAGTCAAGTTTGTAGTGAATGGATCCACCAAGAAAT 4200
QY 4201 GCAATGGAATTAATGCGAGTATATGTAAGCAAGTTGAAAGAAATTTCTAAGAAATTTCT 4260
Db 4201 GCAATGGAATTAATGCGAGTATATGTAAGCAAGTTGAAAGAAATTTCTAAGAAATTTCT 4260
QY 4261 CCCCAAGATCACTGATCACTTTTCAAAAGCTTTTGGCAATACCTCATATGCTTTTAA 4320
Db 4261 CCCCAAGATCACTGATCACTTTTCAAAAGCTTTTGGCAATACCTCATATGCTTTTAA 4320
QY 4321 GTAGAGCTTTCAACCTGAGTGTGAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4380
Db 4321 GTAGAGCTTTCAACCTGAGTGTGAAAGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4380
QY 4381 CCTGAAACAGTTCCAGTGTCCCAAAATATTTGCTTTTCTGATGTTTCAAGTCAACTAGT 4440
Db 4381 CCTGAAACAGTTCCAGTGTCCCAAAATATTTGCTTTTCTGATGTTTCAAGTCAACTAGT 4440
QY 4441 GCAACATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500
Db 4441 GCAACATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500
QY 4501 ACCACTCAACTTCGTTGCTCAAAAGTGGAAATGCGAAATGCGAAATGCGAAATGCGAAAT 4560
Db 4501 ACCACTCAACTTCGTTGCTCAAAAGTGGAAATGCGAAATGCGAAATGCGAAATGCGAAAT 4560
QY 4561 CAAAGAAATTTCAATACCTCTATGAAGTCACTTTAACTGAAGAGACAGTATATGGAATTAAG 4620
Db 4561 CAAAGAAATTTCAATACCTCTATGAAGTCACTTTAACTGAAGAGACAGTATATGGAATTAAG 4620
QY 4621 AAATTTAGTGTATAGATTTCCAGTGTGCGCAATGCTGCGTGTGCGTGTGCGTGTGCGT 4680
Db 4621 AAATTTAGTGTATAGATTTCCAGTGTGCGCAATGCTGCGTGTGCGTGTGCGTGTGCGT 4680
QY 4681 TCAAACTGGAATTTTACAAAGAACTCTGCTGCGCTCCAGATGCTGCTCTCTCTCTCTCTCT 4740
Db 4681 TCAAACTGGAATTTTACAAAGAACTCTGCTGCGCTCCAGATGCTGCTCTCTCTCTCTCTCT 4740
QY 4741 CATGTAGTGAACAACTCACTTTTATGATCATGCAATTAAGCTGAGTGAACCTGCTGCTCAT 4800

4741 CATGTAGTACAACTACACCTTTTAGCATCAGCATAAAGCTGGAGTGAACCTCGTGTCAATT 4800
4801 ACTGGACCAACATGTTTATCTGATTTGATGTCAAATCGGTAGATAATGATGAATTTAATATA 4860
4801 ACTGGACCAACATGTTTATCTGATTTGATGTCAAATCGGTAGATAATGATGAATTTAATATA 4860
4861 TCCTTCATCAAGTCAAATGAAGAAAAATAAACCATAGAAATTAAGATTTAGAAAATATTC 4920
4861 TCCTTCATCAAGTCAAATGAAGAAAAATAAACCATAGAAATTAAGATTTAGAAAATATTC 4920
4921 ACAAGTATCTGTAGTATCACTGCAATTTACCTGGGAAATTTAGTGCTGCAATGATGAGAA 4980
4921 ACAAGTATCTGTAGTATCACTGCAATTTACCTGGGAAATTTAGTGCTGCAATGATGAGAA 4980
4981 GGGAGTCAAGTCTGAAATGATGTTTACTTTTAGAATCAGCCCCAAGGACCCACCT 5040
4981 GGGAGTCAAGTCTGAAATGATGTTTACTTTTAGAATCAGCCCCAAGGACCCACCT 5040
5041 AACACATGACATTTTCAGAAAGATPACAGATGAAGTTTACAAAATTTCAATTAACGTTTCCTT 5100
5041 AACACATGACATTTTCAGAAAGATPACAGATGAAGTTTACAAAATTTCAATTAACGTTTCCTT 5100
5101 CCTCCTTCTCAACTAATGGAATATCCAAAGTATATCAAGCTCTGTTTACCGAGAGAT 5160
5101 CCTCCTTCTCAACTAATGGAATATCCAAAGTATATCAAGCTCTGTTTACCGAGAGAT 5160
5161 GATCCTACTGCTGTCAGATTCACACCTCAGTATTTATACGAAACCAACACANTTCGTC 5220
5161 GATCCTACTGCTGTCAGATTCACACCTCAGTATTTATACGAAACCAACACANTTCGTC 5220
5221 ATTGCAATGCTAGAGGACTAAAAGGTGGACATACATCAATATCAAGTGTTCACGAGTC 5280
5221 ATTGCAATGCTAGAGGACTAAAAGGTGGACATACATCAATATCAAGTGTTCACGAGTC 5280
5281 AATAGTGTGTCAGAGTCCAAAGTTCAGATGAGAAATACCAATGATATCAAGGCTCCA 5340
5281 AATAGTGTGTCAGAGTCCAAAGTTCAGATGAGAAATACCAATGATATCAAGGCTCCA 5340
5341 GCACGACCAAAACCAACCCCTATTTATGATGCCACAGGAAATCTGTTGTGACT 5400
5341 GCACGACCAAAACCAACCCCTATTTATGATGCCACAGGAAATCTGTTGTGACT 5400
5401 TCAACAACTATACAAATCAGAATGCCAATGATGTTACTACAGTATGATCATGACCAATATA 5460
5401 TCAACAACTATACAAATCAGAATGCCAATGATGTTACTACAGTATGATCATGACCAATATA 5460
5461 AAAAAATGTACAGTCTGCGACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAG 5520
5461 AAAAAATGTACAGTCTGCGACAGAAACAGGAGCTCAGCATGATGGAATGTAAACAAAG 5520
5521 TGGTATGATGCATATTTTAAATGAAGCAAGCCATATTTTACAAATGAAGGCTTTCCCTAAC 5580
5521 TGGTATGATGCATATTTTAAATGAAGCAAGCCATATTTTACAAATGAAGGCTTTCCCTAAC 5580
5581 CCTCCATGTACAGAGGAAGACAAAGTTTAGTGCAATGAAGAAATCTACATCATAGGT 5640
5581 CCTCCATGTACAGAGGAAGACAAAGTTTAGTGCAATGAAGAAATCTACATCATAGGT 5640
5641 GCTGATAATGATGCATGATCTCTGGCAATGAAGACAAAATTTGCAATGGAACCACTGAAA 5700
5641 GCTGATAATGATGCATGATCTCTGGCAATGAAGACAAAATTTGCAATGGAACCACTGAAA 5700
5701 CAAAAAAGCAATACCTTTTAAATTTAGAGCTACAAATATTTATGGAACAAATTTACTGAC 5760
5701 CAAAAAAGCAATACCTTTTAAATTTAGAGCTACAAATATTTATGGAACAAATTTACTGAC 5760
5761 TCTGATTTATCTGACCTGTTTAAAGCTTTTAGGGGAAGGACTTTCAGAAAGAACCCGAGAG 5820
5761 TCTGATTTATCTGACCTGTTTAAAGCTTTTAGGGGAAGGACTTTCAGAAAGAACCCGAGAG 5820
5821 ATCATTCTTTCCGTCACTTTGTGATATCTTTTCAATTAATCTCTCTGGAAACAGCTATTTTT 5880
5821 ATCATTCTTTCCGTCACTTTGTGATATCTTTTCAATTAATCTCTCTGGAAACAGCTATTTTT 5880

5881 GCATTTGCAAGAAATTCGACAGAACGAAAGAGGTGGCACATATCTCTCTCAGGATGCA 5940
5881 GCATTTGCAAGAAATTCGACAGAACGAAAGAGGTGGCACATATCTCTCTCAGGATGCA 5940
5941 GAAATTAATGACACTAAATTTGAAGCTGATCAGCTCATCAAGTGGGAGACTGGAACTG 6000
5941 GAAATTAATGACACTAAATTTGAAGCTGATCAGCTCATCAAGTGGGAGACTGGAACTG 6000
6001 AAGGACGAGAGNTTAAACCGGCCAATAAGCGCCAATTCCTCTGCAACATGTTGAAGAG 6060
6001 AAGGACGAGAGNTTAAACCGGCCAATAAGCGCCAATTCCTCTGCAACATGTTGAAGAG 6060
6061 CTTTTCACAAAACAACTAAAGTTTCAAGAAAGAAATTTTCGAAATTAACCAAAATTTCTT 6120
6061 CTTTTCACAAAACAACTAAAGTTTCAAGAAAGAAATTTTCGAAATTAACCAAAATTTCTT 6120
6121 CAGGATCTTTCTTCAACTGATGCTGATCTGCTGGAATGAGCAAAAACCGTTTCCCA 6180
6121 CAGGATCTTTCTTCAACTGATGCTGATCTGCTGGAATGAGCAAAAACCGTTTCCCA 6180
6181 AACATAAAACCATATATAAATAAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTCCA 6240
6181 AACATAAAACCATATATAAATAAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTCCA 6240
6241 GGTTCGGATTTATTTAAATGCCAGCTATATTTCTGCTTATTTATGTCCTCAAAATGAAATTTAT 6300
6241 GGTTCGGATTTATTTAAATGCCAGCTATATTTCTGCTTATTTATGTCCTCAAAATGAAATTTAT 6300
6301 GCTACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACC 6360
6301 GCTACTCAAGGTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGGTGTGGGAAACC 6360
6361 AGGGCAAAAACATTTAGTAAATGCTAAACAGTGTGTTTGAAGAGGAGCGATCAGATGCCAT 6420
6361 AGGGCAAAAACATTTAGTAAATGCTAAACAGTGTGTTTGAAGAGGAGCGATCAGATGCCAT 6420
6421 CAGTATTTGGCCACAGACAAACAGGACGTTACTGCTTTTGGAGATATAGTGAATTAACAAG 6480
6421 CAGTATTTGGCCACAGACAAACAGGACGTTACTGCTTTTGGAGATATAGTGAATTAACAAG 6480
6481 CTAATGGAGGATGTTTCAATAGATTGGACTACAGGATCTGAAATTTGAAAGGATGGG 6540
6481 CTAATGGAGGATGTTTCAATAGATTGGACTACAGGATCTGAAATTTGAAAGGATGGG 6540
6541 GATTCATGACTGTTTGGACAGTGTAACTTTACTGCTGGCCAGAGCATGGGTTCTCTGAG 6600
6541 GATTCATGACTGTTTGGACAGTGTAACTTTACTGCTGGCCAGAGCATGGGTTCTCTGAG 6600
6601 AACAGCGCCCTCTAAATTCACCTTTCTGAAGTTGGTTTCGAGCAAGCAGGGCAATGACACC 6660
6601 AACAGCGCCCTCTAAATTCACCTTTCTGAAGTTGGTTTCGAGCAAGCAGGGCAATGACACC 6660
6661 ACACCTATGATGTTTCACTGCAAGTGTGAGAGTGGTTCGAGCAAGCAGGGCAATGACACC 6720
6661 ACACCTATGATGTTTCACTGCAAGTGTGAGAGTGGTTCGAGCAAGCAGGGCAATGACACC 6720
6721 GACCAATTAACAACATATAAATGACCATGATTTTGGATATATATGCACTAGTGTAGCT 6780
6721 GACCAATTAACAACATATAAATGACCATGATTTTGGATATATATGCACTAGTGTAGCT 6780
6781 GAACTGAGAGGTAAGAAATGTCATGTTGTCAGAAATCTGGCACAGTATATCTTTTACAC 6840
6781 GAACTGAGAGGTAAGAAATGTCATGTTGTCAGAAATCTGGCACAGTATATCTTTTACAC 6840
6841 CAGTGCATTTCTGATCTCTTATCAATTAAGGAAATTAATCAGCCCATCTGTTTGTGTAAC 6900
6841 CAGTGCATTTCTGATCTCTTATCAATTAAGGAAATTAATCAGCCCATCTGTTTGTGTAAC 6900
6901 TATTTCAGCACTTCAGAGATGCACTCTTTTGGAGCGCATGGAGGTGATGTTGAGCTTGAA 6960
6901 TATTTCAGCACTTCAGAGATGCACTCTTTTGGAGCGCATGGAGGTGATGTTGAGCTTGAA 6960

QY 6961 TGGGAGAACCCACTATGTAATATTCAGACCAAGGATACAAATGGAAGAGATTTTAA 7020
Db 6961 TGGGAGAACCCACTATGTAATATTCAGACCAAGGATACAAATGGAAGAGATTTTAA 7020
QY 7021 ATCCAGGGGCCAAAGTTACCCCTCATTTCTCCGAATTCGAACCTTTAAAGAA 7080
Db 7021 ATCCAGGGGCCAAAGTTACCCCTCATTTCTCCGAATTCGAACCTTTAAAGAA 7080
QY 7081 ATATCTATGCTTCTCTCACTGTGCTTT 7108
Db 7081 ATATCTATGCTTCTCTCACTGTGCTTT 7108

RESULT 2
US-08-036-210-21
; Sequence 21, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3430_
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"

US-08-036-210-21

Query Match 36.8%; Score 2618.6; DB 1; Length 3973;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;

QY 4386 AACAGTCCAGGTGTCGCCAATAATATGCTTTTCTGATGTTCACTCACTAGTGCAC 4445
Db 33 AAAAGTTCACAGTGTTCGCCAATAATATGCTTTTCTGATGTTCACTCACTAGTGCAC 92
QY 4446 ATTGACATGATGAAGACCTGACACTATCCTTGCTACTTTCAAATTTACAAATTTACCAC 4505
Db 93 ATTGACATGATGAAGACCTGACACTATCCTTGCTACTTTCAAATTTACAAATTTACCAC 152

QY 4506 TCACTTCTGCTCAAAATGCAAGATGGGAATCCGAAGATGTGTTGAATATCAAAA 4565
Db 153 TCACTTCTGCTCAAAATGCAAGATGGGAATCCGAAGATGTGTTGAATATCAAAA 212
QY 4566 AATTCAATACCTCTATGAAGCTCATTAACTGAAGAGACAGATATATGGATTAAGAAAT 4625
Db 213 AATTCAATACCTCTATGAAGCTCATTAACTGAAGAGACAGATATATGGATTAAGAAAT 272
QY 4626 TAGATGGTATAGATTCGAAGTGGCTGCCAGACCAATGCTGGCTATGGCAATGCTTCAAA 4685
Db 273 TAGATGGTATAGATTCGAAGTGGCTGCCAGACCAATGCTGGCTATGGCAATGCTTCAAA 332
QY 4686 CTGGATTTCTCAAAAACCTCTGCTGGCCCTCCAGATGGTCTCTCTGAAAATGTTCAATGT 4745
Db 333 CTGGATTTCTCAAAAACCTCTGCTGGCCCTCCAGATGGTCTCTCTGAAAATGTTCAATGT 392
QY 4746 AGTAGCAACATCACCTTTTAGCATCAGCATAGCTGGAGTGAACCTGTCTCAATCTGG 4805
Db 393 AGTAGCAACATCACCTTTTAGCATCAGCATAGCTGGAGTGAACCTGTCTCAATCTGG 452
QY 4806 ACCAACAATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAATAATATCTT 4865
Db 453 ACCAACAATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAATAATATCTT 512
QY 4866 CATCAAGTCAAAATGAAGAAAATAAACCATAGAAAATTAAGATTTAGAAAATATTACAAG 4925
Db 513 CATCAAGTCAAAATGAAGAAAATAAACCATAGAAAATTAAGATTTAGAAAATATTACAAG 572
QY 4926 GTATTCTGTAGTATCAGTGTGCTGGAACATTAGTGTGCTGATATGTAGAAAGGAA 4985
Db 573 GTATTCTGTAGTATCAGTGTGCTGGAACATTAGTGTGCTGATATGTAGAAAGGAA 632
QY 4986 GTCAAGTGTGAAATGATTGTTACTACTTTTAGAATCAGCCCAAGAGCCCACTAACAA 5045
Db 633 GTCAAGTGTGAAATGATTGTTACTACTTTTAGAATCAGCCCAAGAGCCCACTAACAA 692
QY 5046 CATGACATTTAGAGATACAGATGAAGTTACAAAATTTCAATTAAGCTTCTCTCTCC 5105
Db 693 CATGACATTTAGAGATACAGATGAAGTTACAAAATTTCAATTAAGCTTCTCTCTCC 752
QY 5106 TTCTCAACCTTAATGGAATATCCAGTATATCAAGCTCTGCTTTTACCGAGAGATGATCC 5165
Db 753 TTCTCAACCTTAATGGAATATCCAGTATATCAAGCTCTGCTTTTACCGAGAGATGATCC 812
QY 5166 TACTGCTGTCCAGATTCACAACTCAGTATTTATACAGAAAACCAACACATTCGTCAATGC 5225
Db 813 TACTGCTGTCCAGATTCACAACTCAGTATTTATACAGAAAACCAACACATTCGTCAATGC 872
QY 5226 AATGTAGAAAGGACTAAAAGGTGGACATACATCAATATCAGTGTGTTACGAGTCAATAG 5285
Db 873 AATGTAGAAAGGACTAAAAGGTGGACATACATCAATATCAGTGTGTTACGAGTCAATAG 932
QY 5286 TGCTGGTGCAGGTCCAAAGGTTCCGATGAGATACCAATGATATCAAAAGCTCCAGCAG 5345
Db 933 TGCTGGTGCAGGTCCAAAGGTTCCGATGAGATACCAATGATATCAAAAGCTCCAGCAG 992
QY 5346 ACCAAAAACCAACCAACCCCTATTATTATGATGCCACAGGAAAATGCTTGTGACTTCAAC 5405
Db 993 ACCAAAAACCAACCAACCCCTATTATTATGATGCCACAGGAAAATGCTTGTGACTTCAAC 1052
QY 5406 AACAAATTAATCAGATGCCAATATGTTTACAGTATGATGATGAGCAATATAAAAAA 5465
Db 1053 AACAAATTAATCAGATGCCAATATGTTTACAGTATGATGATGAGCAATATAAAAAA 1112
QY 5466 TGTAAGAGTCTTGCGACAGAAACAGGAGCTCAGCATGATGAGAAATGTAAACAAAGTGA 5525
Db 1113 TGTAAGAGTCTTGCGACAGAAACAGGAGCTCAGCATGATGAGAAATGTAAACAAAGTGA 1172
QY 5526 TGATGATATTTTAAATGAAGGCCATATTTTAAATGAAGGCTTTCTTAAACCTCC 5585
Db 1173 TGATGATATTTTAAATGAAGGCCATATTTTAAATGAAGGCTTTCTTAAACCTCC 1232
QY 5586 ATGTAAGAGGAAAGACAAAGTTTTAGTGGCAATGAAGAAATCTACATCATATGAGTCTGA 5645

Db 1233 ATGTACAGAGGAAACACAAAGTTTAGTGCATGAAGAAATCTACATCATAGGTGCTGA 1292
Qy 5646 TAATGATGATGATTCCTGGCAATGAAGACAAATTTGCAATGGACCACTGAACCAAA 5705
Db 1293 TAATGATGATGATTCCTGGCAATGAAGACAAATTTGCAATGGACCACTGAACCAAA 1352
Qy 5706 AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGCAATTTACTGACTCGA 5765
Db 1353 AAAGCAATCTTATTTAAATTTAGAGCTACAAATATTTAGGCAATTTACTGACTCGA 1412
Qy 5766 TTATTTCTGACCTGTTAAAGACTTTAGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCAT 5825
Db 1413 TTATTTCTGACCTGTTAAAGACTTTAGGGAAGGACTTTTCAGAAAGAACCGTAGAGATCAT 1472
Qy 5826 TCTTTCGGTCACTTTGATGCTCTTCAATAATCTCTCTGGAACAGCTATTTTGGCAT 5885
Db 1473 TCTTTCGGTCACTTTGATGCTCTTCAATAATCTCTCTGGAACAGCTATTTTGGCAT 1532
Qy 5886 TGCAAGAAATTCGACAGAGCAGAAAGAGTGSCACATCTCTCTCAGGATGCGAAAT 5945
Db 1533 TGCAAGAAATTCGACAGAGCAGAAAGAGTGSCACATCTCTCTCAGGATGCGAAAT 1592
Qy 5946 TATTTGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTTGGAACTGAAGGA 6005
Db 1593 TATTTGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTTGGAACTGAAGGA 1652
Qy 6006 CGAGAGATTAACCGG----- 6020
Db 1653 CGAGAGATTAACCGGATCTCTTCAATTTCTTTAGACGCAAGGAGATTTTGTCTATCCA 1712
Qy 6021 -----GCCAATAAGCAAGAAATCTCTCTCAGCAATGT 6053
Db 1713 GTTACTTAGTTAAGAAATCCATCAAGCCAAATAGCAAGAAATCTCTCTGCAATGT 1772
Qy 6054 TGAAGAGCTTTGCAACAAACCAACCTAAAGTTTCAAGAAAGATTTTTCGGAATTAACAAA 6113
Db 1773 TGAAGAGCTTTGCAACAAACCAACCTAAAGTTTCAAGAAAGATTTTTCGGAATTAACAAA 1832
Qy 6114 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCG 6173
Db 1833 ATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCG 1892
Qy 6174 TTTCCCAACATAAAACCATATAATAATACAGAGTAAAGCTGATAGCTGACGCTAG 6233
Db 1893 CTTCACCAACATAAAACCAT- --ATAATATAACAGATAAGCTGATAGCTGACGCTAG 1949
Qy 6234 TGTTCCAGGTTCCGGATTATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGA 6293
Db 1950 TGTTCCAGGTTCCGGATTATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGA 2009
Qy 6294 ATTTATTTGCTACTCAAGGTTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGTGTG 6353
Db 2010 ATTTATTTGCTACTCAAGGTTCCACTACAGGAAACAGTTGGAGATTTTGGAGAAATGTGTG 2069
Qy 6354 GGAACCGGGCAAAACATTAAGTATGCTTAACACAGTGTTTTGAAGAGCGGATCAG 6413
Db 2070 GGAACCGAGAGCAAAACATTAAGTATGCTTAACACAGTGTTTTGAAGAGCGGATCAG 2129
Qy 6414 ATGCCATCAGTATTTGGCCAGAGCAACAGCCAGTACTGTCTTTGGAGATATAGTAT 6473
Db 2130 ATGCCATCAGTATTTGGCCAGAGCAACAGCCAGTACTGTCTTTGGAGATATAGTAT 2189
Qy 6474 TACAAAGCTAATGGAGATGTTCAAAATAGATTGGAATATCAGGAGATCTGAAAATGAAAG 6533
Db 2190 TACAAAGCTAATGGAGATGTTCAAAATAGATTGGAATATCAGGAGATCTGAAAATGAAAG 2249
Qy 6534 GCATGGGATGATGATCTGTTCCAGAGTGAATCTTCTGCTGGCCAGAGATGGGT 6593
Db 2250 GCATGGGATGATGATCTGTTCCAGAGTGAATCTTCTGCTGGCCAGAGATGGGT 2309
Qy 6594 TCTGTGAACAGCGCCCTCTAAATTCATTTGTGAAGTTGGTTTCGAGCAGAGCGGCACA 6653

Db 2310 TCTGTGAACAGCGCCCTCTAAATTCATTTGTGAAGTTGGTTTCGAGCAAGCAGGCACA 2369
Qy 6654 TGACACCAACCACTATGATTTGTTCACTGCAAGTCTGCAAGTCTGGAAGCTTGGAGTTTAT 6713
Db 2370 TGACACCAACCACTATGATTTGTTCACTGCAAGTCTGCAAGTCTGGAAGCTTGGAGTTTAT 2429
Qy 6714 TCTCTGTGACCACTTTAAACACCAACATATAAATGACCATGATTTTGTGGATATATATGGACT 6773
Db 2430 TCTCTGTGACCACTTTAAACACCAACATATAAATGACCATGATTTTGTGGATATATATGGACT 2489
Qy 6774 AGTAGCTGAACGAGAGTGAAGAAATGTGCAATGTGCAAGTCTGGAAGTCTGGCAGATATCTTT 6833
Db 2490 AGTAGCTGAACGAGAGTGAAGAAATGTGCAATGTGCAAGTCTGGCAGATATCTTT 2549
Qy 6834 TTTTACACAGTGCATTTCTGATCTCTTATCAATAAGGGAAGTAAATCAGCCCATCTGTTT 6893
Db 2550 TTTTACACAGTGCATTTCTGATCTCTTATCAATAAGGGAAGTAAATCAGCCCATCTGTTT 2609
Qy 6894 TGTAACTATTCAGCACTTTCAAGAGATGGACTCTTTTGGACGCCATGGAAGTGTATGTTGA 6953
Db 2610 TGTAACTATTCAGCACTTTCAAGAGATGGACTCTTTTGGACGCCATGGAAGTGTATGTTGA 2669
Qy 6954 GCTTGAATGGGAAGAAACCACTATGTAATATTCAGACCAAAAGGATACAAATTTGGAAGAGA 7013
Db 2670 GCTTGAATGGGAAGAAACCACTATGTAATATTCAGACCAAAAGGATACAAATTTGGAAGAGA 2729
Qy 7014 TTTTAAATCCCAGGGGCCAAAGTTTACCCCTCATTTCCGAATTTGAAATGTGCAACCT 7073
Db 2730 TTTTAAATCCCAGGGGCCAAAGTTTACCCCTCATTTCCGAATTTGAAATGTGCAACCT 2789
Qy 7074 TAAAGAAATATCTATGCTTTCTCTCACTGTGCTTT 7108
Db 2790 TAAAGAAATATCTATGCTTTCTCTCACTGTGCTTT 2824

RESULT 3

US-08-449-609-21
; Sequence 21, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 24-MAY-1995
; APPLICATION NUMBER: US/08/449,609
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TBLEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3430
; OTHER INFORMATION: /note= "N=x=unknown nucleotide"
; US-08-449-609-21

Query Match 36.8%; Score 2618.6; DB 2; Length 3973;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;

QY 4386 AACAGTTCACAGTTCACCAAAATATGCTTTCTGATGTTTCAGTCAACTAGTCAAC 4445
DB 33 AAAAGTTCACAGTTCACCAAAATATGCTTTCTGATGTTTCAGTCAACTAGTCAAC 92

QY 4446 ATTGACATGATGAAGACCTGACACTATCTCTGGCTACTTTCAAAATTTACAAATTTACCAC 4505
DB 93 ATTGACATGATGAAGACCTGACACTATCTCTGGCTACTTTCAAAATTTACAAATTTACCAC 152

QY 4506 TCACTTCGTGCTCAAAATGAAAGTGAAGATCGGAATCGGAATGTTGAATATCAAAA 4565
DB 153 TCACTTCGTGCTCAAAATGAAAGTGAAGATCGGAATGTTGAATATCAAAA 212

QY 4566 AATTCATACCTCTATGAAGCTCACTTAACTGAAGACAGTATATGATTAAGAATTT 4625
DB 213 AATTCATACCTCTATGAAGCTCACTTAACTGAAGACAGTATATGATTAAGAATTT 272

QY 4626 TAGATGATATAGATTCACAGTGGCTGCCAGCAATGCTGGCTATGGCAATGCTTCAAA 4685
DB 273 TAGATGATATAGATTCACAGTGGCTGCCAGCAATGCTGGCTATGGCAATGCTTCAAA 332

QY 4686 CTGGATTTCTCAAAATCTGCTGCCCTCCAGATGCTCTCTGAAATGTTCAATGT 4745
DB 333 CTGGATTTCTCAAAATCTGCTGCCCTCCAGATGCTCTCTGAAATGTTCAATGT 392

QY 4746 AGTAGCAATCACTTTAGCATCAGCATAGCTGGAGTGAACCTGCTCAATTAAGT 4805
DB 393 AGTAGCAATCACTTTAGCATCAGCATAGCTGGAGTGAACCTGCTCAATTAAGT 452

QY 4806 ACCAATGTTATCTGATGATGTCAAATCGGTAGATATGATGAATTTAATATATCTCT 4865
DB 453 ACCAATGTTATCTGATGATGTCAAATCGGTAGATATGATGAATTTAATATATCTCT 512

QY 4866 CATCAAGTCAAAATGAAGAAATAAACCATAGAAATTTAAGATTTAGAAATATTCACAA 4925
DB 513 CATCAAGTCAAAATGAAGAAATAAACCATAGAAATTTAAGATTTAGAAATATTCACAA 572

QY 4926 GTATCTGTAGTCACTGCTCAATTTACTGGACATTAAGTCTGCTATGATAGAGGAA 4985
DB 573 GTATCTGTAGTCACTGCTCAATTTACTGGACATTAAGTCTGCTATGATAGAGGAA 632

QY 4986 GTCAAGTGTGAATGATGTTACTCTTTAGATCAGCCCAAGACCCACTTAACAA 5045
DB 633 GTCAAGTGTGAATGATGTTACTCTTTAGATCAGCCCAAGACCCACTTAACAA 692

QY 5046 CATGACATTTCAAGATACCATGATGAAGTTACAAATTTCAATTAACGTTCTCTCTCC 5105
DB 693 CATGACATTTCAAGATACCATGATGAAGTTACAAATTTCAATTAACGTTCTCTCTCC 752

QY 5106 TTCTCAACTTAATGGAATATCAAGTATATCAAGCTCTGTTTACCGAAGATGATCC 5165
DB 753 TTCTCAACTTAATGGAATATCAAGTATATCAAGCTCTGTTTACCGAAGATGATCC 812

QY 5166 TACTGCTGCTCCAGATTCACAACTCAGTATATACAGAAACCAACACATTCGCTATTC 5225
DB 813 TACTGCTGCTCCAGATTCACAACTCAGTATATACAGAAACCAACACATTCGCTATTC 872

QY 5226 AATGCTAGAGGAGTAAAGGTGGACATACATACATATCAGTGTGTACGCACTCAATAG 5285
DB 873 AATGCTAGAGGAGTAAAGGTGGACATACATACATATCAGTGTGTGTACGCACTCAATAG 932

QY 5286 TGCTGTGTCAGGTCCAAAGTTCCTGATGAGATTAACCATGATATCAAGCTCCAGCAGC 5345
DB 933 TGCTGTGTCAGGTCCAAAGTTCCTGATGAGATTAACCATGATATCAAGCTCCAGCAGC 992

QY 5346 ACCAAAAACCAACCCCTATTTATGATGCCACAGGAAAACTCTCTGTGTGCTTCAAC 5405
DB 993 ACCAAAAACCAACCCCTATTTATGATGCCACAGGAAAACTCTCTGTGTGCTTCAAC 1052

QY 5406 AACCAATTACATCAGATGCAATATGTTACTACATGATGATCATGAGCAATTAACAA 5465
DB 1053 AACCAATTACATCAGATGCAATATGTTACTACATGATGATCATGAGCAATTAACAA 1112

QY 5466 TGTACAAAGTGTCTGCGACAGAAACAGGAGCTCAGCATGATGGAATTTGAACAAAGTGGTA 5525
DB 1113 TGTACAAAGTGTCTGCGACAGAAACAGGAGCTCAGCATGATGGAATTTGAACAAAGTGGTA 1172

QY 5526 TGAATGATATTTTAAATGAAGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCC 5585
DB 1173 TGAATGATATTTTAAATGAAGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCC 1232

QY 5586 ATGTACAGAGGAAACAGCAAAAGTTTGTAGTGCATGAAGAAATCTACATCATAGTGTCTGA 5645
DB 1233 ATGTACAGAGGAAACAGCAAAAGTTTGTAGTGCATGAAGAAATCTACATCATAGTGTCTGA 1292

QY 5646 TAAATGATCATGATTTCTGCGCAATGAAGCAAAATTTGCAATGGACCACTGAACCCCAA 5705
DB 1293 TAAATGATCATGATTTCTGCGCAATGAAGCAAAATTTGCAATGGACCACTGAACCCCAA 1352

QY 5706 AAGCAATATCTTATTTAAATTTAGAGCTACAAATATTTAGGGAATTTTACTGACTCTGA 5765
DB 1353 AAGCAATATCTTATTTAAATTTAGAGCTACAAATATTTAGGGAATTTTACTGACTCTGA 1412

QY 5766 TTATTTCTGACCTGTTTAAAGCTTTAGGGAAGAGCACTTTCAGAAAGAACCGTAGAGATCAT 5825
DB 1413 TTATTTCTGACCTGTTTAAAGCTTTAGGGAAGAGCACTTTCAGAAAGAACCGTAGAGATCAT 1472

QY 5826 TCTTTTCGTCACCTTTGTGATCTCTTCAATTAATCTCTTGGAAACAGCTATTTTTCAT 5885
DB 1473 TCTTTTCGTCACCTTTGTGATCTCTTCAATTAATCTCTTGGAAACAGCTATTTTTCAT 1532

QY 5886 TGCAGAAATTTGACAGAGCAAGAGAGGTGGCACTACTCTCTCAGGATTCAGAAAT 5945
DB 1533 TGCAGAAATTTGACAGAGCAAGAGAGGTGGCACTACTCTCTCAGGATTCAGAAAT 1592

QY 5946 TATTTGACATTAATTTGAAGCTGATCAGCTCATCAGTGGGAGAGCTGGAAGTGAAGGA 6005
DB 1593 TATTTGACATTAATTTGAAGCTGATCAGCTCATCAGTGGGAGAGCTGGAAGTGAAGGA 1652

QY 6006 CGAGAGATTAACGCG----- 6020
DB 1653 CGAGAGATTAACGCGATCTCTTCTTTTGTAGCGAAGAGATTTTGTCTATCCA 1712

QY 6021 -----GCCAATTAAGCAAGAAATCTCTCTGCAACATGT 6053
DB 1713 GTTACTTACTTATAGAAATCCATCAAGCCAAATAGCAAGAAATCTCTCTGCAACATGT 1772

QY 6054 TGAAGAGCTTTGACAAACAAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTCACAAA 6113
DB 1773 TGAAGAGCTTTGACAAACAAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTCACAAA 1832

QY 6114 ATTTCTTCAGGATCTTTCTTCACTGATGCTGCTCTGGAATAGAGCAAAACCG 6173
DB 1833 ATTTCTTCAGGATCTTTCTTCACTGATGCTGCTCTGGAATAGAGCAAAACCG 1892

QY 6174 TTTTCCCAACATATAAACCACTATAATAATTAACAGATTAAGCTGATAGTGCAGCTAG 6233
DB 1893 CTTTCCCAACATATAAACCACTATAATAATTAACAGATTAAGCTGATAGTGCAGCTAG 1949

QY 6234 TGTTCAGGTTCCGATTTATTAATGSCCAGCTATATTTCTGCTTATTTATGTCCTAATGA 6293

Db 1950 TGTTCAGGTTGGATTATTAATGCCAGTATATTTCTGGTATTATATGTCCTCAATGA 2009
Qy 6294 ATTATATGCTACTCAAGTCCACTACACAGAACAGTTGGAGATTTTGGAGAAATGTTGG 6353
Db 2010 ATTATATGCTACTCAAGTCCACTACACAGAACAGTTGGAGATTTTGGAGAAATGTTGG 2069
Qy 6354 GGAACACAGGGCAAAACATTAGTAATGCTAACACAGTGTTTTGAAGAGGACGGATCAG 6413
Db 2070 GGAACACAGAGCAAAACATTAGTAATGCTAACACAGTGTTTTGAAGAGGACGGATCAG 2129
Qy 6414 ATGCCATCAGTATGTCGACAGGACCAACAGGACGTTACTGTCTTTGGAGATATAGTGAT 6473
Db 2130 ATGCCATCAGTATGTCGACAGGACCAACAGGACGTTACTGTCTTTGGAGATATAGTGAT 2189
Qy 6474 TACAAAGCTAATGGAGGATGTTCAAAATAGATTTGGACTATCAGGATCTGAAAATTGAAG 6533
Db 2190 TACAAAGCTAATGGAGGATGTTCAAAATAGATTTGGACTATCAGGATCTGAAAATTGAAG 2249
Qy 6534 GCATGGGATTCATGACTGTTTCGACAGTGAATCTTACTGCTGGCCAGACATGGGT 6593
Db 2250 GCATGGGATTCATGACTGTTTCGACAGTGAATCTTACTGCTGGCCAGACATGGGT 2309
Qy 6594 TCCAGAGACAGCGCCCTCTAATTCATCTTTGTGAAGTTGGTTTCGAGCAAGCAGGACACA 6653
Db 2310 TCCAGAGACAGCGCCCTCTAATTCATCTTTGTGAAGTTGGTTTCGAGCAAGCAGGACACA 2369
Qy 6654 TGACACCAACCTATGATTTGTCACAGTGTCTGAGTGTGGAGTTGGAAGAACTGGAGTTTAT 6713
Db 2370 TGACACCAACCTATGATTTGTCACAGTGTCTGAGTGTGGAGTTGGAAGAACTGGAGTTTAT 2429
Qy 6714 TGCCTGGACCAATTAACACACATATTAATGACCATGATTTTGTGGATATATATGACT 6773
Db 2430 TGCCTGGACCAATTAACACACATATTAATGACCATGATTTTGTGGATATATATGACT 2489
Qy 6774 AGTAGCTGAACCTGAGAAAGTGAAGATGTGCATGTGTCAGAAATCTGGCAACATATCTT 6833
Db 2490 AGTAGCTGAACCTGAGAAAGTGAAGATGTGCATGTGTCAGAAATCTGGCAACATATCTT 2549
Qy 6834 TTTACACCAAGTGAATCTGATCTCTTATCAATAAGGAAAGTAATCAGCCCATCTGTTT 6893
Db 2550 TTTACACCAAGTGAATCTGATCTCTTATCAATAAGGAAAGTAATCAGCCCATCTGTTT 2609
Qy 6894 TGTAACTATTCAGCACTTCAGAGATGCACTCTTTGGACGCCATGGAAGTGTGTTGA 6953
Db 2610 TGTAACTATTCAGCACTTCAGAGATGCACTCTTTGGACGCCATGGAAGTGTGTTGA 2669
Qy 6954 GCTTGAATGGGAAGAACCACTATGTAAATATTCAGACCAAGGATACAAATTTGGAAGAGA 7013
Db 2670 GCTTGAATGGGAAGAACCACTATGTAAATATTCAGACCAAGGATACAAATTTGGAAGAGA 2729
Qy 7014 TTTTAAATCCAGGGGCAAGTTAACCCCTCATTTCTCCGAATGAAATGTGCAACT 7073
Db 2730 TTTTAAATCCAGGGGCAAGTTAACCCCTCATTTCTCCGAATGAAATGTGCAACT 2789
Qy 7074 TAAAGAAATATCTATGCTTCTCTCACTGTCCTTT 7108
Db 2790 TAAAGAAATATCTATGCTTCTCTCACTGTCCTTT 2824

RESULT 4
US-09-361-096A-21
; Sequence 21, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609

; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 3973
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2
; NAME/KEY: modified base
; LOCATION: (1)..(3973)
; OTHER INFORMATION: n = unknown nucleotide
; NAME/KEY: CDS
; LOCATION: (1)..(2694)
; US-09-361-096A-21

Query Match 36.8%; Score 2618.6; DB 4; Length 3973;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;
Qy 4386 AACAGTCCCAGTGTCCCAAAATATTGCTTTTCTGATGTTCACTAGTCAACTAGTGCAC 4445
Db 33 AAAAGTCCCAGTGTCCCAAAATATTGCTTTTCTGATGTTCACTAGTCAACTAGTGCAC 92
Qy 4446 ATTGACATGGATAGACCTGACATCTCTTGGCTACTTTCAAAATTACAAAATTACCAC 4505
Db 93 ATTGACATGGATAGACCTGACATCTCTTGGCTACTTTCAAAATTACAAAATTACCAC 152
Qy 4506 TCAACTCTGCTCAAAAATGCGAAATGCGAAATGCGAAATGCGAAATGCGAAATGCGAA 4565
Db 153 TCAACTCTGCTCAAAAATGCGAAATGCGAAATGCGAAATGCGAAATGCGAAATGCGAA 212
Qy 4566 AATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGAATTAAGAAAT 4625
Db 213 AATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGAATTAAGAAAT 272
Qy 4626 TAGATGTTATAGATTTCCAGTGGCTGCCAGACCAATGCTGGCTATGCGAATGCTTCAA 4685
Db 273 TAGATGTTATAGATTTCCAGTGGCTGCCAGACCAATGCTGGCTATGCGAATGCTTCAA 332
Qy 4686 CTGGATTTCTACAAAACTCTGCTGGCTCCAGATGCTCTCTGAAAAATGTTCACTGT 4745
Db 333 CTGGATTTCTACAAAACTCTGCTGGCTCCAGATGCTCTCTGAAAAATGTTCACTGT 392
Qy 4746 AGTAGCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGTCACTACTGG 4805
Db 393 AGTAGCAACATCACTTTTAGCATCAGCATAGCTGGAGTGAACCTGCTGTCACTACTGG 452
Qy 4806 ACCAATGTTTCTGATGATGTCAAATCGGTAGATTAATGATTAATATATATCTT 4865
Db 453 ACCAATGTTTCTGATGATGTCAAATCGGTAGATTAATGATTAATATATATCTT 512
Qy 4866 CATCAAGTCAAAATGAAGAAAAATAAACCATAGAAATTAAGAAATTAAGAAATTAAGAA 4925
Db 513 CATCAAGTCAAAATGAAGAAAAATAAACCATAGAAATTAAGAAATTAAGAAATTAAGAA 572
Qy 4926 GTATTCTGTAGTATCCTGCAATTTACTGGGAACATTAAGTGTGCTGATATGTAAGAGGAA 4985
Db 573 GTATTCTGTAGTATCCTGCAATTTACTGGGAACATTAAGTGTGCTGATATGTAAGAGGAA 632
Qy 4986 GTCAAGTGTGAATGATTTGTTACTTACTTTAGAAATCAGCCCCCAAGGACCCACCTAACAA 5045
Db 633 GTCAAGTGTGAATGATTTGTTACTTACTTTAGAAATCAGCCCCCAAGGACCCACCTAACAA 692
Qy 5046 CATGACATTTTCAGAGATACCAAGATGAAGTTTACAAAATTTCAATTTAACTCTCTCTCC 5105
Db 693 CATGACATTTTCAGAGATACCAAGATGAAGTTTACAAAATTTCAATTTAACTCTCTCTCC 752
Qy 5106 TTCTCAACCTAATGGAATATCCAAATATCAAGCTCTGGTTTACCGAAGAGATGATCC 5165
Db 753 TTCTCAACCTAATGGAATATCCAAATATCAAGCTCTGGTTTACCGAAGAGATGATCC 812

Qy	5166	TACTGCTGTCAGATTCACAACCTTCAGTATTATACAGAAACCAACACATTCGTGTCATGTC	5225
Db	813	TACTGCTGTCAGATTCACAACCTTCAGTATTATACAGAAACCAACACATTCGTGTCATGTC	872
Qy	5226	AATGCTAGAGGACTTAAAGGTGACATACATACAATATCAGTGTGTTTACCGAGTCAATAG	5285
Db	873	AATGCTAGAGGACTTAAAGGTGACATACATACAATATCAGTGTGTTTACCGAGTCAATAG	932
Qy	5286	TGCTGGTCAGGTCCAAAGTTCGAGTGAAGTAACCATGGATATCAAAAGCTCCAGCACG	5345
Db	933	TGCTGGTCAGGTCCAAAGTTCGAGTGAAGTAACCATGGATATCAAAAGCTCCAGCACG	992
Qy	5346	ACCAAAAACCAACCAACCCCTATTATTATGATGCCACAGGAAAACCTGCTGTGACTTCAAC	5405
Db	993	ACCAAAAACCAACCAACCCCTATTATTATGATGCCACAGGAAAACCTGCTGTGACTTCAAC	1052
Qy	5406	AACAATTTACAATCAGAATGCCAATATGTTACTACAGTGATGATCATCGACCAATAAAAAA	5465
Db	1053	AACAATTTACAATCAGAATGCCAATATGTTACTACAGTGATGATCATCGACCAATAAAAAA	1112
Qy	5466	TGTAACAAGTCTTCGACAGAAACAGAGCTCAGCATGATGGAAATGTAAACAAAGTGGTA	5525
Db	1113	TGTAACAAGTCTTCGACAGAAACAGAGCTCAGCATGATGGAAATGTAAACAAAGTGGTA	1172
Qy	5526	TGATGCATATTTTAAATAAGCAGGCCATATTTTACAANTGAAGGCTTTTCTTAACCCCTCC	5585
Db	1173	TGATGCATATTTTAAATAAGCAGGCCATATTTTACAANTGAAGGCTTTTCTTAACCCCTCC	1232
Qy	5586	ATGTACAGAAGGAAAGCAAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGTGTCTGA	5645
Db	1233	ATGTACAGAAGGAAAGCAAAAGTTTAGTGGCAATGAAGAAATCTACATCATAGTGTCTGA	1292
Qy	5646	TAATGCATGCATGATTCCTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAACCAAA	5705
Db	1293	TAATGCATGCATGATTCCTGGCAATGAAGCAAAATTTGCAATGGACCACTGAAACCAAA	1352
Qy	5706	AAAGCATACTTATTTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGA	5765
Db	1353	AAAGCATACTTATTTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGA	1412
Qy	5766	TTATTCTGACCTGTTAAGACTTTAGGGGAAGGACTTTCAGAAAGAACCGTAGAGATCAT	5825
Db	1413	TTATTCTGACCTGTTAAGACTTTAGGGGAAGGACTTTCAGAAAGAACCGTAGAGATCAT	1472
Qy	5826	TCCTTCGGTCACCTTGTGTATCCTTTTCAAATAATTCCTCTTGGAACAGCTATTTTGCATT	5885
Db	1473	TCCTTCGGTCACCTTGTGTATCCTTTTCAAATAATTCCTCTTGGAACAGCTATTTTGCATT	1532
Qy	5886	TGCAGGAATTCGACAGAGCAGAAAGGTGGCAATCTCTCTCTCAGAGATCGAGAAAT	5945
Db	1533	TGCAGGAATTCGACAGAGCAGAAAGGTGGCAATCTCTCTCTCAGAGATCGAGAAAT	1592
Qy	5946	TATTGACATTAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTGAAGGA	6005
Db	1593	TATTGACATTAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTGAAGGA	1652
Qy	6006	CGAGAGATTTAACGGC-----	6020
Db	1653	CGAGAGATTTAACGGC-----	1712
Qy	6021	-----GCCAATAAGCAAGAAATCCTTCTCTCGCAACATGT	6053
Db	1713	GTTACTTAGTTATAGAAAATTCGATCAAGCCCAATAAGCAAGAAATCCTTCTCTCGCAACATGT	1772
Qy	6054	TGAAGAGCTTTGCACAAAACCAACCTTAAAGTTTCAAGAGAAATTTTTCGGAATTCACAAA	6113
Db	1773	TGAAGAGCTTTGCACAAAACCAACCTTAAAGTTTCAAGAGAAATTTTTCGGAATTCACAAA	1832
Qy	6114	ATTTCTTCAGGAATCTTTCTTCCTCAACTGATGCTGATCTGCTCTGGAAATAGACAAAAACCG	6173
Db	1833	ATTTCTTCAGGAATCTTTCTTCCTCAACTGATGCTGATCTGCTCTGGAAATAGACAAAAACCG	1892

RESULT 5
US-08-036-210-14
; Sequence 14, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P. H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,609
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,210
FILING DATE: 23-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2692 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2149
OTHER INFORMATION: /note= "N=x-unknown nucleotide"
US-08-449-609-14

Query Match 15.7%; Score 1118.8; DB 2; Length 2692;
Best Local Similarity 93.7%; Pred. No. 6.6e-289;
Matches 1218; Conservative 0; Mismatches 7; Indels 75; Gaps 2;
QY 5881 GCATTTGCAAGATTTCGACAGAGCAGAGAGAGGTTGGCACAATCTCTCTCAGGATGCA 5940
DB 247 GCTTGTGTTTGAATTCGACAGAGCAGAGAGAGGTTGGCACAATCTCTCTCAGGATGCA 306
QY 5941 GAAATTAATGACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTG 6000
DB 307 GAAATTAATGACACTAAATTTGAAGCTGGATCAGCTCATCAGTGGCAGACCTGGAACTG 366
QY 6001 AAGGACGAGATTACGCG----- 6020
DB 367 AAGGACGAGATTACGCGGATCTCTTCATTTTCTTTAGACGCAAGGAGATTTTGTGTC 426
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTCGCAA 6048
DB 427 ATCCAGTTACTAGTATAGAAATCCATCAAGCCATTAAGCAAGAAATCTCTCTCGAA 486
QY 6049 CATGTTGAAGAGCTTTGCAACAAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTA 6108
DB 487 CATGTTGAAGAGCTTTGCAACAAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTA 546
QY 6109 CCMAATTTCTCAGGATCTTTCTCACTGATGCTGATCTGCTTTGGAATAGACAAA 6168
DB 547 CCAAAATTTCTCAGGATCTTTCTCACTGATGCTGATCTGCTTTGGAATAGACAAA 606
QY 6169 AACCGTTTCCAAACATAAACCATATAATAATAACAGAGTAAAGCTGATGCTGAC 6228
DB 607 AACCGTTTCCAAACATAAACCAT---ATAATAACAGAGTAAAGCTGATGCTGAC 663
QY 6229 GCTAGTGTTCAGGTTCCGATTAATAATGCGAGCTATATTTCTGGTATTATATGCTCA 6288
DB 664 GCTAGTGTTCAGGTTCCGATTAATAATGCGAGCTATATTTCTGGTATTATATGCTCA 723

QY 6289 AATGAATTTATTGCTACTCAAGGTCCACTACACAGGAAACAGTTGGAGATTTTGGAGAAATG 6348
DB 724 AATGAATTTATTGCTACTCAAGGTCCACTACACAGGAAACAGTTGGAGATTTTGGAGAAATG 783
QY 6349 GTGTGGAAACAGGCGCAAAACATTTAGTAATGCTAAACAGAGTGTGTTTGAAGAAGCAGCG 6408
DB 784 GTGTGGAAACAGGCGCAAAACATTTAGTAATGCTAAACAGAGTGTGTTTGAAGAAGCAGCG 843
QY 6409 ATCAGATGCCATCAGTATTGGCCAGAGGACAAACAGCCAGTTACTCTCTTTGGAGATATA 6468
DB 844 ATCAGATGCCATCAGTATTGGCCAGAGGACAAACAGCCAGTTACTCTCTTTGGAGATATA 903
QY 6469 GTGATTACAAAGCTAATGAGGATGTTCAAAATAGATTGGAATCAAGGATCTGAAAAATT 6528
DB 904 GTGATTACAAAGCTAATGAGGATGTTCAAAATAGATTGGAATCAAGGATCTGAAAAATT 963
QY 6529 GAAAGGCATGGGATTCGATGATGTTTCGACAGTGTAACTTTACTCTGCTGGCCAGAGCAT 6588
DB 964 GAAAGGCATGGGATTCGATGATGTTTCGACAGTGTAACTTTACTCTGCTGGCCAGAGCAT 1023
QY 6589 GGGGTTCTTGAGAACAGCGCCCTCTAAATTTCACTTTGTGAAGTTGGTTTGAGCAAGCAGG 6648
DB 1024 GGGGTTCTTGAGAACAGCGCCCTCTAAATTTCACTTTGTGAAGTTGGTTTGAGCAAGCAGG 1083
QY 6649 GCACATGACACCACTATGATTTGTTCACTGCGAGTGTGGAGTTGGAAGAACTGGAGTT 6708
DB 1084 GCACATGACACCACTATGATTTGTTCACTGCGAGTGTGGAGTTGGAAGAACTGGAGTT 1143
QY 6709 TTTATTGCTCTGAGCCATTTAAACAAATATAAATGACATGATTTTGTGGATATATAT 6768
DB 1144 TTTATTGCTCTGAGCCATTTAAACAAATATAAATGACATGATTTTGTGGATATATAT 1203
QY 6769 GGACTAGTAGCTGAACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGATCTGGCAGATAT 6828
DB 1204 GGACTAGTAGCTGAACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGATCTGGCAGATAT 1263
QY 6829 ATCTTTTACACAGTGCATTTCTGGATCTCTTATCAAAATAAGGAAAGTAAATCAGCCCATC 6888
DB 1264 ATCTTTTACACAGTGCATTTCTGGATCTCTTATCAAAATAAGGAAAGTAAATCAGCCCATC 1323
QY 6889 TGTTTTGTTAACTATTTCAGCACTTCAGAAAGATGGAATCTTTTGGACGCCATGGAAGTGAT 6948
DB 1324 TGTTTTGTTAACTATTTCAGCACTTCAGAAAGATGGAATCTTTTGGACGCCATGGAAGTGAT 1383
QY 6949 GTTGAGCTTGAATGGAGAACCACTATGTAATAATTCAGACCAAGAGTAAATCAATTGGA 7008
DB 1384 GTTGAGCTTGAATGGAGAACCACTATGTAATAATTCAGACCAAGAGTAAATCAATTGGA 1443
QY 7009 AGAGATTTTAAATCCAGGGCCAAAGTTACCCCTCATTTTCCGAATTTGAAATGTGC 7068
DB 1444 AGAGATTTTAAATCCAGGGCCAAAGTTACCCCTCATTTTCCGAATTTGAAATGTGC 7108
QY 7069 AACCTTAAAGAAATATCTATGCTTTCTCTCATCTGTGCTTTT 7108
DB 1504 AACCTTAAAGAAATATCTATGCTTTCTCTCATCTGTGCTTTT 1543

RESULT 7

US-09-361-096A-14
; Sequence 14, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210

;; PRIOR FILING DATE: 1995-03-23
;; NUMBER OF SEQ ID NOS: 53
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 14
;; LENGTH: 2692

;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: modified base
;; LOCATION: (1)..(2692)
;; OTHER INFORMATION: n = unknown nucleotide
;; NAME/KEY: CDS
;; LOCATION: (92..139, 259..1414)
US-09-361-096A-14

Query Match 15.7%; Score 1118.8; DB 4; Length 2692;

Best Local Similarity 93.7%; Pred. No. 6.6e-289;
Matches 1218; Conservative 0; Mismatches 7; Indels 75; Gaps 2;

QY	5881	GCATTTGCAAGAAATTCGACAGAGCAGAAAGAGGTGGCACATACCTCTCTCAGGATGCA	5940
DB	247	GCTTGTTTGAATTCGACAGAGCAGAAAGAGGTGGCACATACCTCTCTCAGGATGCA	306
QY	5941	GAATTAATTGACACTAAATTTGAAGCTGGATCAGCTCATCACTGTCGACAGCTGGAAC	6000
DB	307	GAATTAATTGACACTAAATTTGAAGCTGGATCAGCTCATCACTGTCGACAGCTGGAAC	366
QY	6001	AAGGACGAGAGATTAACGGG-----	6020
DB	367	AAGGACGAGAGATTAACGGGATACTCTTCATTTTCTTTAGACGCAAGGAGATTTTGTG	426
QY	6021	-----GCCAATAAGCAAGAAATCCATCAAGCCCAATGAAGAAATCTCTCTGCAA	6048
DB	427	ATCCAGTTACTAGTTATAGAAATCCATCAAGCCCAATGAAGAAATCTCTCTGCAA	486
QY	6049	CATGTTGAAGAGCTTTGCAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTA	6108
DB	487	CATGTTGAAGAGCTTTGCAACAACTAAAGTTTCAAGAGAAATTTTCGGAATTA	546
QY	6109	CCAAATTTCTCAGATCTTTCTCAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	6168
DB	547	CCAAATTTCTCAGATCTTTCTCAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	606
QY	6169	AACGGTTTCCCAACATAAAACCATATAATAATAACAGAGTAAAGCTAGTAGCTAC	6228
DB	607	AACGGTTTCCCAACATAAAACCATATAATAATAACAGAGTAAAGCTAGTAGCTAC	663
QY	6229	GCTAGTGTTCAGGTTCCGATTTATATTAATGCCAGCTATATTTCTGTTTATTTATGTC	6288
DB	664	GCTAGTGTTCAGGTTCCGATTTATATTAATGCCAGCTATATTTCTGTTTATTTATGTC	723
QY	6289	AATGAATTTATGCTACTCAAGTCCACTACAGGAACTAGTTGGAGATTTTGGAGATG	6348
DB	724	AATGAATTTATGCTACTCAAGTCCACTACAGGAACTAGTTGGAGATTTTGGAGATG	783
QY	6349	GTGTGGAAACACAGGCAAAACATTTAGTAACTGCTTAACACACTGTTTGAAGAGCAG	6408
DB	784	GTGTGGAAACACAGGCAAAACATTTAGTAACTGCTTAACACACTGTTTGAAGAGCAG	843
QY	6409	ATCAGATGCCATCAGTATTGGCCAGAGGCAACAGCCAGTTACTGCTTTGGAGATATA	6468
DB	844	ATCAGATGCCATCAGTATTGGCCAGAGGCAACAGCCAGTTACTGCTTTGGAGATATA	903
QY	6469	GTGATTAACAACTAATGAGGATGTTCAAAATAGATTTGAGCTATCAGGGATCTGAAAT	6528
DB	904	GTGATTAACAACTAATGAGGATGTTCAAAATAGATTTGAGCTATCAGGGATCTGAAAT	963
QY	6529	GAAAGCATGGGATTTGATGACTGTTCCACAGTGTAACTTTACTGCTGCTGGCAGAGAT	6588
DB	964	GAAAGCATGGGATTTGATGACTGTTCCACAGTGTAACTTTACTGCTGCTGGCAGAGAT	1023
QY	6589	GGGGTTCCTGAGAAACAGCGCCCTCTAAATTTCACTTTGTGAAGTTGGTTCGAGCAAGC	6648

DB	1024	GGGGTTCCTGAGAACAGCGCCCTCTAAATTCATTTGTGAAGTTGGTTCGAGCAAGCAGG	1083
QY	6649	GCACATGACACACACCTATGATTTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	6708
DB	1084	GCACATGACACACACCTATGATTTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1143
QY	6709	TTTATTGCTCTGACCATTTTAAACACACATATAAATGACCATGATTTTGTGGATATAT	6768
DB	1144	TTTATTGCTCTGACCATTTTAAACACACATATAAATGACCATGATTTTGTGGATATAT	1203
QY	6769	GGACTAGTAGCTGAACTGAGAAAGTGAAGAGTGCATGCTGCTGCTGCTGCTGCTGCTG	6828
DB	1204	GGACTAGTAGCTGAACTGAGAAAGTGAAGAGTGCATGCTGCTGCTGCTGCTGCTGCTG	1263
QY	6829	ATCTTTTACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	6888
DB	1264	ATCTTTTACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1323
QY	6889	TGTTTGTGTTAACTATTCAGCACTTCAGAAAGTGGACTCTTTTGGACGCCATGGAAGGT	6948
DB	1324	TGTTTGTGTTAACTATTCAGCACTTCAGAAAGTGGACTCTTTTGGACGCCATGGAAGGT	1383
QY	6949	GTGAGCTTGAATGGGAGAAACCACTATGTAAATATTCAGACCAAAAGGATCAATTTGA	7008
DB	1384	GTGAGCTTGAATGGGAGAAACCACTATGTAAATATTCAGACCAAAAGGATCAATTTGA	1443
QY	7009	AGAGATTTTAAATCCAGGGCCAAAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7068
DB	1444	AGAGATTTTAAATCCAGGGCCAAAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	7108
QY	7069	AACCTTAAAGAAATATCTATGCTTCTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTG	1503
DB	1504	AACCTTAAAGAAATATCTATGCTTCTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTG	1543

RESULT 8

US-08-036-210-10
; Sequence 10, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
US-08-036-210-10

Query Match      15.1%; Score 1071.6; DB 1; Length 2309;
Best Local Similarity 99.4%; Pred. No. 2.7e-276;
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 6015 AACGGGCCAATAGCAAGAAATCTTCTCGCAACATGTTGAAGAGCTTTGCAACAA 6074
DB 70 AATGAGGCCAATAGCAAGAAATCTTCTCGCAACATGTTGAAGAGCTTTGCAACAA 129
QY 6075 CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATACCAGAAATTTCTTCAGATCTTTCTTC 6134
DB 130 CAACCTAAAGTTTCAAGAGAAATTTTCGGAAATACCAGAAATTTCTTCAGATCTTTCTTC 189
QY 6135 AACTGATGCTGATCTGCTTGGAAATAGACCAAAACCGTTTCCCAAAACATAAAACATA 6194
DB 190 AACTGATGCTGATCTGCTTGGAAATAGACCAAAACCGTTTCCCAAAACATAAAACAT- 248
QY 6195 TAATAATAATAACAGAGTAAGCTGATGCTGACGCTAGTGTTCAGAGTTTCGAGTTATAT 6254
DB 249 --ATAATAATAACAGAGTAAGCTGATGCTGACGCTAGTGTTCAGAGTTTCGAGTTATAT 306
QY 6255 TAATGCCAGCTATATTTCTGTTTATTTATGTCGAAATGAATTTATTTGCTACTCAAGGTC 6314
DB 307 TAATGCCAGCTATATTTCTGTTTATTTATGTCGAAATGAATTTATTTGCTACTCAAGGTC 366
QY 6315 ACTTACCAGGACAGTTTGGAGATTTTGGAGAAATGTTGGGAAACAGGCAACCAAAACATT 6374
DB 367 ACTTACCAGGACAGTTTGGAGATTTTGGAGAAATGTTGGGAAACAGGCAACCAAAACATT 426
QY 6375 AGTAATGCTAAACACAGTGTTTTGAAGAGGACGATCAGATGCTATGCTATTTGGCCAGA 6434
DB 427 AGTAATGCTAAACACAGTGTTTTGAAGAGGACGATCAGATGCTATGCTATTTGGCCAGA 486
QY 6435 GGACAAACAGCCAGTGTCTTTGGAGATATAGTGAATTAACAGCTAATGAGAGATGT 6494
DB 487 GGACAAACAGCCAGTGTCTTTGGAGATATAGTGAATTAACAGCTAATGAGAGATGT 546
QY 6495 TCAATATAGATTGCACTATCAGGATCTGAAATTTGAAGGCAATGGGATTTGCTGACTGT 5554
DB 547 TCAATATAGATTGCACTATCAGGATCTGAAATTTGAAGGCAATGGGATTTGCTGACTGT 606
QY 6555 TCGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGTTCTGTGAGAACAGGCGCCCTCT 6614
DB 607 TCGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGTTCTGTGAGAACAGGCGCCCTCT 666
QY 6615 AATTCACTTTGGAAGTTGGTTTCAGCAAGAGGCGACATGACACACACCTATGATGT 6674
DB 667 AATTCACTTTGGAAGTTGGTTTCAGCAAGAGGCGACATGACACACACCTATGATGT 726
QY 6675 TCACTGCTGCTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGGACATTTAACA 6734
DB 727 TCACTGCTGCTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGGACATTTAACA 786
QY 6735 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAAGTCA 6794
DB 787 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAAGTCA 846
QY 6795 AAGAATGTGATGGTGCAGAAATCTGGCACAGTATATCTTTTACACAGTGAATTTGGA 6854
DB 847 AAGAATGTGATGGTGCAGAAATCTGGCACAGTATATCTTTTACACAGTGAATTTGGA 906
QY 6855 TCTCTTATCAATAAGGGAAGTAATCAGGCCATCTGTTTGTAACTATTTCAGACTTCA 6914
DB 907 TCTCTTATCAATAAGGGAAGTAATCAGGCCATCTGTTTGTAACTATTTCAGACTTCA 966
```

```
QY 6915 GAAGATGGAGTCTTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAACCCAC 6974
DB 967 GAAGATGGAGTCTTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAACCCAC 1026
QY 6975 TATGTAATAATTTACAGACCAAGGATACAAATTCGAGAGATTTTAAATCCAGGGGCCAA 7034
DB 1027 TATGTAATAATTTACAGACCAAGGATACAAATTCGAGAGATTTTAAATCCAGGGGCCAA 1086
QY 7035 AGTTACCCCTCATCTTCCGAATTTGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT 7094
DB 1087 AGTTACCCCTCATCTTCCGAATTTGAAATGTGCAACCTTAAAGAAATATCTATGCTTCT 1146
QY 7095 CTCACGTGCGCTTT 7108
DB 1147 CTCACGTGCGCTTT 1160
```

RESULT 9

```
US-08-449-609-10
; Sequence 10, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSES: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; NAME: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
US-08-449-609-10
```

```
Query Match      15.1%; Score 1071.6; DB 2; Length 2309;
Best Local Similarity 99.4%; Pred. No. 2.7e-276;
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;
```

```
QY 6015 AACGGGCCAATAAGCAAGAAATCCTTCTCTGCAACATGTTGAAGAGCTTTGACAAACAA 6074
|
|
|
Db 70 AATGAGGCCAATAAGCAAGAAATCCTTCTCTGCAACATGTTGAAGAGCTTTGACAAACAA 129
|
|
|
QY 6075 CAACCTAAAGTTTCAAGAAGAATTTTCGGAATTAACAAATTTCTTCAGGATCTTCTTC 6134
|
|
|
Db 130 CAACCTAAAGTTTCAAGAAGAATTTTCGGAATTAACAAATTTCTTCAGGATCTTCTTC 189
|
|
|
QY 6135 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAACATAAAACCAT 6194
|
|
|
Db 190 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAACATAAAACCAT 248
|
|
|
QY 6195 TAATAATAAACAAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGAGTTGCGATTATAT 6254
|
|
|
Db 249 --ATAATAAACAAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGAGTTGCGATTATAT 306
|
|
|
QY 6255 TAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTTGCTACTCAAGTCC 6314
|
|
|
Db 307 TAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTTGCTACTCAAGTCC 366
|
|
|
QY 6315 ACTTACCAGAAACAGTTGGAGATTTTGGAGATGTTGCGGAATGTTGCGGAACCCAGGCAAAACATT 6374
|
|
|
Db 367 ACTTACCAGAAACAGTTGGAGATTTTGGAGATGTTGCGGAACCCAGGCAAAACATT 426
|
|
|
QY 6375 AGTAATGCTTAAACAACAGTGTGTTGAAAAGGAGCGGATCAGATGCCATCAGTATTTGCCAGA 6434
|
|
|
Db 427 AGTAATGCTTAAACAACAGTGTGTTGAAAAGGAGCGGATCAGATGCCATCAGTATTTGCCAGA 486
|
|
|
QY 6435 GGACAAACAAGCCAGTTACTGTTTGGAGATATAGTATTGAAAGCTAATCGAGGATGT 6494
|
|
|
Db 487 GGACAAACAAGCCAGTTACTGTTTGGAGATATAGTATTGAAAGCTAATCGAGGATGT 546
|
|
|
QY 6495 TCAATATAGATTGGAATCTCAGGATCTGAAATTTGAAAGGATGCGGATTCGATGCTGT 6554
|
|
|
Db 547 TCAATATAGATTGGAATCTCAGGATCTGAAATTTGAAAGGATGCGGATTCGATGCTGT 606
|
|
|
QY 6555 TCGACAGTGTAACTTTACTGCTGCGCAGAGATGGGTTCTCTGAGAAACAGCGCCCTCT 6614
|
|
|
Db 607 TCGACAGTGTAACTTTACTGCTGCGCAGAGATGGGTTCTCTGAGAAACAGCGCCCTCT 666
|
|
|
QY 6615 AATTCACTTTGTGAAGTTGGTTGAGCAAGCAGGCGACATGACACACACCTATGATGT 6674
|
|
|
Db 667 AATTCACTTTGTGAAGTTGGTTGAGCAAGCAGGCGACATGACACACACCTATGATGT 726
|
|
|
QY 6675 TCACCTGCTGCTGGAGTTGGAGACTGAGTTTATTTATGCTCTGACCAATTAACACA 6734
|
|
|
Db 727 TCACCTGCTGCTGGAGTTGGAGAACTGGAGTTTATTTGCTCTGACCAATTAACACA 786
|
|
|
QY 6735 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAACTGAGAGTGA 6794
|
|
|
Db 787 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGCTGAACTGAGAGTGA 846
|
|
|
QY 6795 AAGAAATGTCATGGTGCAGAAATCTGSCACAGTATATCTTTTACACCACTGCAATCTGGA 6854
|
|
|
Db 847 AAGAAATGTCATGGTGCAGAAATCTGSCACAGTATATCTTTTACACCACTGCAATCTGGA 906
|
|
|
QY 6855 TCTCTTATCAATAAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTGACACTTCA 6914
|
|
|
Db 907 TCTCTTATCAATAAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTGACACTTCA 966
|
|
|
QY 6915 GAAGATGGAATCTTTTGGACCGCAATGAAGTGTGAGCTTGAATGGGAAGAACAC 6974
|
|
|
Db 967 GAAGATGGAATCTTTTGGACCGCAATGAAGTGTGAGCTTGAATGGGAAGAACAC 1026
|
|
|
QY 6975 TATGTAAATATTCAGACCAAGGATACAAATGGGAAGAGATTTTAAATCCAGGGGCCAA 7034
|
|
|
Db 1027 TATGTAAATATTCAGACCAAGGATACAAATGGGAAGAGATTTTAAATCCAGGGGCCAA 1086
|
|
|
QY 7035 AGTTACCCCTCATTTCTCGGAATTTGAATGTGCAACCTTAAGAAATATCTATGCTTCT 7094
|
|
|
Db 1087 AGTTACCCCTCATTTCTCGGAATTTGAATGTGCAACCTTAAGAAATATCTATGCTTCT 1146
|
|
|
QY 7095 CTCACCTGTGCTTT 7108
```

```
Db 1147 CTCACCTGTGCTTT 1160
|
|
|
RESULT 10
US-09-361-096A-10
; Sequence 10, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 2309
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31D
; NAME/KEY: modified base
; LOCATION: (1)..(2309)
; OTHER INFORMATION: n = unknown nucleotide
; NAME/KEY: CDS
; LOCATION: (65)..(1030)
US-09-361-096A-10

Query Match 15.1%; Score 1071.6; DB 4; Length 2309;
Best Local Similarity 99.4%; Pred. No. 2.7e-276;
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 6015 AACGGGCCAATAAGCAAGAAATCCTTCTCTGCAACATGTTGAAGAGCTTTGACAAACAA 6074
|
|
|
Db 70 AATGAGGCCAATAAGCAAGAAATCCTTCTCTGCAACATGTTGAAGAGCTTTGACAAACAA 129
|
|
|
QY 6075 CAACCTAAAGTTTCAAGAAGAATTTTCGGAATTAACAAATTTCTTCAGGATCTTCTTC 6134
|
|
|
Db 130 CAACCTAAAGTTTCAAGAAGAATTTTCGGAATTAACAAATTTCTTCAGGATCTTCTTC 189
|
|
|
QY 6135 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAACATAAAACCAT 6194
|
|
|
Db 190 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAACATAAAACCAT 248
|
|
|
QY 6195 TAATAATAAACAAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGGATCTTTCAGGATCT 6254
|
|
|
Db 249 --ATAATAAACAAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGGATCTTTCAGGATCT 306
|
|
|
QY 6255 TAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTTGCTACTCAAGTCC 6314
|
|
|
Db 307 TAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTTGCTACTCAAGTCC 366
|
|
|
QY 6315 ACTTACCAGAAACAGTTGGAGATTTTGGAGATGTTGCGGAATGTTGCGGAACCCAGGCAAAACATT 6374
|
|
|
Db 367 ACTTACCAGAAACAGTTGGAGATTTTGGAGATGTTGCGGAACCCAGGCAAAACATT 426
|
|
|
QY 6375 AGTAATGCTTAAACAACAGTGTGTTGAAAAGGAGCGGATCAGATGCCATCAGTATTTGCCAGA 6434
|
|
|
Db 427 AGTAATGCTTAAACAACAGTGTGTTGAAAAGGAGCGGATCAGATGCCATCAGTATTTGCCAGA 486
|
|
|
QY 6435 GGACAAACAAGCCAGTTACTGTTTGGAGATATAGTATTGAAAGCTAATCGAGGATGT 6494
|
|
|
Db 487 GGACAAACAAGCCAGTTACTGTTTGGAGATATAGTATTGAAAGCTAATCGAGGATGT 546
|
|
|
QY 6495 TCAATATAGATTGGAATCTCAGGATCTGAAATTTGAAAGGATGCGGATTCGATGCTGT 6554
|
|
|
```

Db 547 TCAATAGATTGGACTATCAGGATCTGAAATTTGAAAGCATGGGATTGCATGACTGT 606
Qy 6555 TCGACAGTGTAACTTTACTGCTGGCCAGAGCATGGGGTTCTTGAGAACAGCCCTCT 6614
Db 607 TCGACAGTGTAACTTTACTGCTGGCCAGAGCATGGGGTTCTTGAGAACAGCCCTCT 666
Qy 6615 AATTCATTTGTGAGTGGTTCGAGCAAGCAGGCGACATGACACACACCTATGATTGT 6674
Db 667 AATTCATTTGTGAGTGGTTCGAGCAAGCAGGCGACATGACACACACCTATGATTGT 726
Qy 6675 TCACTGAGTGGTGGAGTTCGAGCAAGCAGGCGACATGACACACACCTATGATTGT 6734
Db 727 TCACTGAGTGGTGGAGTTCGAGCAAGCAGGCGACATGACACACACCTATGATTGT 786
Qy 6735 ACATATAATGACCATGATTTTGTGATATATATGACATAGTAGTACCTGAACTGAGAGTGA 6794
Db 787 ACATATAATGACCATGATTTTGTGATATATATGACATAGTAGTACCTGAACTGAGAGTGA 846
Qy 6795 ARGAGTGCATGGTGCAGATCTGGCAGATATATCTTTTACACAGTGCATTCTGGA 6854
Db 847 ARGAGTGCATGGTGCAGATCTGGCAGATATATCTTTTACACAGTGCATTCTGGA 906
Qy 6855 TCTCTTATCAATAAGGAGTAAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCACTTCA 6914
Db 907 TCTCTTATCAATAAGGAGTAAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCACTTCA 966
Qy 6915 GAAGATGGACTCTTTTGGACCCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAACCCAC 6974
Db 967 GAAGATGGACTCTTTTGGACCCCATGGAAGGTGATGTTGAGCTTGAATGGGAAGAACCCAC 1026
Qy 6975 TATGTAATAATTCAGACCAAGATCAATGGAAGATTTTAAATCCAGGGGCCAA 7034
Db 1027 TATGTAATAATTCAGACCAAGATCAATGGAAGATTTTAAATCCAGGGGCCAA 1086
Qy 7035 AGTTACCCCTCAATCTTCCGAATGGAATGTCGAACCTTAAAGAAATATCTATGCTTCT 7094
Db 1087 AGTTACCCCTCAATCTTCCGAATGGAATGTCGAACCTTAAAGAAATATCTATGCTTCT 1146
Qy 7095 CTCACGTGCTTT 7108
Db 1147 CTCACGTGCTTT 1160

RESULT 11

US-08-036-210-6

; Sequence 6, Application US/08036210

; Patent No. 5585233

; GENERAL INFORMATION:

; APPLICANT: Moller, Niels P.H.

; APPLICANT: Moller, Karin B.

; APPLICANT: Ullrich, Axel

; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE

; TITLE OF INVENTION: PHOSPHATASE

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDWARDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/036,210

; FILING DATE: 23-MAR-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mibrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1630
; OTHER INFORMATION: /note= "N=x=unknown nucleotide"
; US-08-036-210-6

Query Match 11.1%; Score 789.6; DB 1; Length 2173;
Best Local Similarity 86.9%; Pred. No. 9.8e-201;
Matches 951; Conservative 0; Mismatches 4; Indels 139; Gaps 2;
Qy 6015 AACGGCGCCCAATAAGCAAGAAATCTTCTGCAACATGTTGAAGAGCTTTGCAACAACAA 6074
Db 70 AATGAGGCCAATAAGCAAGAAATCTTCTGCAACATGTTGAAGAGCTTTGCAACAACAA 129
Qy 6075 CAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 6134
Db 130 CAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 189
Qy 6135 AACTGATGCTGATCTGCCTTGGAAATAGAGCAAAAAACCGTTTCCCAAAACATAAAACCATTA 6194
Db 190 AACTGATGCTGATCTGCCTTGGAAATAGAGCAAAAAACCGTTTCCCAAAACATAAAACCAT- 248
Qy 6195 TAATAATAATACACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGAGTTCCGATTATAT 6254
Db 249 --ATAATAATACACAGATTAAGCTGATAGCTGACGCTAGTGTTCAGAGTTCCGATTATAT 306
Qy 6255 TAATGCCAGCTATATTCTTGGTTATTTATGTCCAAATGAATTTATTTGCTACTCAAGGTCC 6314
Db 307 TAATGCCAGCTATATTCTTGGTTATTTATGTCCAAATGAATTTATTTGCTACTCAAGGTCC 366
Qy 6315 ACTACCAAGAACAGTTGGAGATTTTGGAGATTTGGTGGGAAACAGGGCAAAAAACATT 6374
Db 367 ACTACCAAGAACAGTTGGAGATTTTGGAGATTTGGTGGGAAACAGGGCAAAAAACATT 426
Qy 6375 AGTAATGCTAACACAGTGTTTTGAAGAGGAGCGATCAGATGCCATCAGTATTGGCCAGA 6434
Db 427 AGTAATGCTAACACAGTGTTTTGAAGAGGAGCGATCAGATGCCATCAGTATTGGCCAGA 486
Qy 6435 GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGAATTAACAAGCTAATGGAGGATGT 6494
Db 487 GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGAATTAACAAGCTAATGGAGGATGT 546
Qy 6495 TCAATATAGATTGGACTATCAGGGATCTGAAATTTGAAAGGCATGGGGATTTGCATGCTGT 6554
Db 547 TCAATATAGATTGGACTATCAGGGATCTGAAATTTGAAAGGCATGGGGATTTGCATGCTGT 606
Qy 6555 TCGACAGTGTAACTTTTACTGCTGCGCCAGAGCATGGGGTTCTTGAGAACAGCGCCCTCT 6614
Db 607 TCGACAGTGTAACTTTTACTGCTGCGCCAGAGCATGGGGTTCTTGAGAACAGCGCCCTCT 666
Qy 6615 AATTCACTTTTGAAGTTGGTTTCGAGCAAGCAGGGGCACATGACACCAACCTATGATTGT 6674
Db 667 AATTCACTTTTGAAGTTGGTTTCGAGCAAGCAGGGGCACATGACACCAACCTATGATTGT 726
Qy 6675 TCACCTGAGTGTGGAGTTGGAAGAACCTGGAGTTTTTTTATTTGCTCTGGAACCACTTAACACA 6734
Db 727 TCACCTGCA----- 734
Qy 6735 ACATATAATGACCATGATTTTGTGATATATATGGACTAGTAGCTGAACTGGAAGTGA 6794
Db 735 ----- 734

QY 6795 AAGAATGTGCATGGTGCAGAACTCGGCACAGTATATCTTTTACACAGTGCATCTGGA 6854
Db |||||||-----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 6855 TCTCTTATCAATAAAGGGAAGTAATCAGCCATCTGTTTGTAACTATTCAGCACTTCA 6914
Db |||||||-----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 6915 GAAGATGGACTCTTTTGGACCCATGGAAGTGATGTTGAGCTTGAATGGGAAGAACCCAC 6974
Db |||||||-----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 6975 TATGTAATAATTCAGACCAAGATACAAATGGGAAGATTTTAAATCCCGGGGCCAA 7034
Db |||||||-----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 7035 AGTTACCCCTCATCTTCGGAATGGAATGTCACCTTAAAGAAATATCTATGCTTCT 7094
Db |||||||-----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 7095 CTCACGTGCTCTTT 7108
Db |||||||-----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 1011 CTCACGTGCTCTTT 1024

RESULT 12

US-08-449-609-6
; Sequence 6, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mierock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: 1630
; OTHER INFORMATION: /note= "N=x=unknown nucleotide"
US-08-449-609-6
Query Match 11.1%; Score 789.6; DB 2; Length 2173;
Best Local Similarity 86.9%; Pred. No. 9.8e-201;
Matches 951; Conservative 0; Mismatches 4; Indels 139; Gaps 2;
QY 6015 AACGGGCAATAAGCAAGAAATCTCTCTGCAACATGTTGAAGAGCTTTTGACAAACAA 6074
Db 70 AATGAGGCCAATAAGCAAGAAATCTCTCTGCAACATGTTGAAGAGCTTTTGACAAACAA 129
QY 6075 CAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 6134
Db 130 CAACCTAAAGTTTCAAGAAAGAAATTTTCGGAATTTACCAAAATTTCTTCAGGATCTTTCTTC 189
QY 6135 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCGTTTCCCAAAATTAACCAACATA 6194
Db 190 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAACCGTTTCCCAAAATTAACCAACATA 248
QY 6195 TAATAATAATACACAGTAAGCTCATAGCTGACGCTAGTGTTCAGAGTTTCGGATTATAT 6254
Db 249 --ATAATAATAACAGATTAAGCTCATAGCTGACGCTAGTGTTCAGAGTTTCGGATTATAT 306
QY 6255 TAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATTTATGCTACTCAAGGTCC 6314
Db 307 TAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATTTATGCTACTCAAGGTCC 366
QY 6315 ACTACAGCAAGTTGGAGATTTTGGAGATGTTGGGAAACAGGGCAAAACATTT 6374
Db 367 ACTACAGCAAGTTGGAGATTTTGGAGATGTTGGGAAACAGGGCAAAACATTT 426
QY 6375 AGTAATGCTAACACAGTGTGTTTGAAGAGGAGGATCAGATGCCATCAGTATTTGGCCAGA 6434
Db 427 AGTAATGCTAACACAGTGTGTTTGAAGAGGAGGATCAGATGCCATCAGTATTTGGCCAGA 486
QY 6435 GGACAAACAGCCAGTACTGCTCTTTGGAGATATAGTGATTAACAAGCTTAATCGAGGATGT 6494
Db 487 GGACAAACAGCCAGTACTGCTCTTTGGAGATATAGTGATTAACAAGCTTAATCGAGGATGT 546
QY 6495 TCAAAATAGATTGGCACTATCAGGGATCTGAAATTTGAAGGCAATGGGATTTGATGACTGT 6554
Db 547 TCAAAATAGATTGGCACTATCAGGGATCTGAAATTTGAAGGCAATGGGATTTGATGACTGT 606
QY 6555 TCGACAGTGTAACTTTTACTGCTGCCAGAGCATGGGTTCTGAGAAACAGGCCCTCT 6614
Db 607 TCGACAGTGTAACTTTTACTGCTGCCAGAGCATGGGTTCTGAGAAACAGGCCCTCT 666
QY 6615 AATTCACTTTGTGAAGTTGGTTTCGAGCAAGCGGCAATGACACACACCTATGATGT 6674
Db 667 AATTCACTTTGTGAAGTTGGTTTCGAGCAAGCGGCAATGACACACACCTATGATGT 726
QY 6675 TCACCTGAGTGTGGAGTTGGAAAGTGGAGTTTATTTGCTCTGGACCATTTAACACA 6734
Db 727 TCACCTGCA----- 734
QY 6735 ACATATAAATGACCATGATTTTGTGATATATATGAGCTAGTAGCTGAAGTGAAGTGA 6794
Db 735 ----- 734
QY 6795 AAGAATGTGCATGGTGCAGAAATCTGGCAAGTATATCTTTTACACAGTGCATCTGGA 6854
Db 735 -----GGCAGTATATCTTTTACACAGTGCATCTGGA 770
QY 6855 TCTCTTATCAATAAAGGGAAGTAATCAGCCATCTGTTTGTAACTATTCAGCACTTCA 6914
Db 771 TCTCTTATCAATAAAGGGAAGTAATCAGCCCACTCTGTTTGTAACTATTCAGCACTTCA 830
QY 6915 GAAGATGGACTCTTTTGGACCCATGGAAGTGATGTTGAGCTTGAATGGGAAGAAACCCAC 6974
Db 831 GAAGATGGACTCTTTTGGACCCATGGAAGTGATGTTGAGCTTGAATGGGAAGAAACCCAC 890
QY 6975 TATGTAATAATTCAGACCAAGGATACAAATGGGAAGATTTTAAATCCCGGGGCCAA 7034

```
Db 891 TAATGTAATATTCAGACCAAGGATCAATTTGGAAGAGATTTTAAATCCAGGGCCAA 950
Qy 7035 AGTTACCCCTCATTTCTTCGAAATGGAATGTGCAACCTTTAAAGAAATATCTATGCTTCT 7094
Db 951 AGTTACCCCTCATTTCTTCGAAATGGAATGTGCAACCTTTAAAGAAATATCTATGCTTCT 1010
Qy 7095 CTCAGTGTGCTTT 7108
Db 1011 CTCAGTGTGCTTT 1024

RESULT 13
US-09-361-096A-6
; Sequence 6, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2173
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31C
; NAME/KEY: modified base
; LOCATION: (1)..(1630)
; OTHER INFORMATION: n = unknown nucleotide
; NAME/KEY: CDS
; LOCATION: (65)..(1066)
US-09-361-096A-6

Query Match 11.1%; Score 789.6; DB 4; Length 2173;
Best Local Similarity 86.9%; Pred. No. 9.8e-201;
Matches 951; Conservative 0; Mismatches 4; Indels 139; Gaps 2;

Qy 6015 AACGGGCCAATTAAGCAAGAAATCCCTTCCTGCAACATGTTGAAGAGTTTGCACAAACAA 6074
Db 70 AATGAGGCCAATTAAGCAAGAAATCCCTTCCTGCAACATGTTGAAGAGTTTGCACAAACAA 129
Qy 6075 CAACCTAAAGTTTCAAGAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTTCTTC 6134
Db 130 CAACCTAAAGTTTCAAGAGAAATTTTCGGAATACCAAAATTTCTTCAGGATCTTTCTTC 189
Qy 6135 AACTGATGCTGATCTGCTTGGAAATAGACAAACAAACCGTTTCCCAACATAAAACATA 6194
Db 190 AACTGATGCTGATCTGCTTGGAAATAGACAAACAAACCGTTTCCCAACATAAAACAT- 248
Qy 6195 TAATAATAATAACAGAGTAAGCTGATGATGCTGACGCTAGTGTTCAGGTTTCGGATTATAT 6254
Db 249 --ATAATAATAACAGAGTAAGCTGATGATGCTGACGCTAGTGTTCAGGTTTCGGATTATAT 306
Qy 6255 TAATGCCAGCTATTTCTGTTTATTTATGTCGCAATGAAATTTATTTGCTACTCAAGTCC 6314
Db 307 TAATGCCAGCTATTTCTGTTTATTTATGTCGCAATGAAATTTATTTGCTACTCAAGTCC 366
Qy 6315 ACTACACAGGAACAGTTGGAGATTTTGGGAATGGTGGGAACCGAGGCAAAACATTT 6374
Db 367 ACTACACAGGAACAGTTGGAGATTTTGGGAATGGTGGGAACCGAGGCAAAACATTT 426
Qy 6375 AGTAATGCTAACACAGTGTTTTGAAGAAAGGACCGGATCAGATGCCATCAGTATTGGCCAGA 6434
```

```
Db 427 AGTAATGCTAACACACAGTGTTTTGAAGAAAGGACGGATCAGATGCATCAGTATTGGCCAGA 486
Qy 6435 GGAACAACAAGCCAGTTACTGTCTTTGGAGATATAGTATTAACAAGCTAATGAGGATGT 6494
Db 487 GGAACAACAAGCCAGTTACTGTCTTTGGAGATATAGTATTAACAAGCTAATGAGGATGT 546
Qy 6495 TCAATAGATTGGAGTCTCAGGGATCTGAAATTTGAAAGCATGGGATTTGCATGACTGT 6554
Db 547 TCAATAGATTGGAGTCTCAGGGATCTGAAATTTGAAAGCATGGGATTTGCATGACTGT 606
Qy 6555 TCGACAGTGTAACTTTTACTGCTGCGCAGAGCATGGGTTCTTGAGAAACAGCGCCCTCT 6614
Db 607 TCGACAGTGTAACTTTTACTGCTGCGCAGAGCATGGGTTCTTGAGAAACAGCGCCCTCT 666
Qy 6615 AATTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGCGACATGACACCACTATGATTGT 6674
Db 667 AATTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGCGACATGACACCACTATGATTGT 726
Qy 6675 TCACTGAGTGTGGAGTTGGAAAGAACTGGAGTTTATTTGCTCTGACCAATTAACACA 6734
Db 727 TCACTGCA----- 734
Qy 6735 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAGTGAAGTGA 6794
Db 735 ----- 734
Qy 6795 AAGAATGTGATGGTGCAGAAATCTGGCAGACAGTATATCTTTTACACAGTGCATTCTGGA 6854
Db 735 -----GSCACAGTATATCTTTTACACAGTGCATTCTGGA 770
Qy 6855 TCTTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCATTTCA 6914
Db 771 TCTTTATCAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTGTTAACTATTTCAGCATTTCA 830
Qy 6915 GAAGATGGACTCTTTGAGCGCCATGCAAGTGTGATGTTGAGCTTGAATGGGAAGAACCCAC 6974
Db 831 GAAGATGGACTCTTTGAGCGCCATGCAAGTGTGATGTTGAGCTTGAATGGGAAGAACCCAC 890
Qy 6975 TATGTAATATTCAGACCAAGGATCAATTTGGAAGAGATTTTAAATCCAGGGGCCAA 7034
Db 891 TATGTAATATTCAGACCAAGGATCAATTTGGAAGAGATTTTAAATCCAGGGGCCAA 950
Qy 7035 AGTTACCCCTCATTTCTCGAATTTGGAATTTGCAACCTTTAAAGAAATATCTATGCTTCT 7094
Db 951 AGTTACCCCTCATTTCTCGAATTTGGAATTTGCAACCTTTAAAGAAATATCTATGCTTCT 1010
Qy 7095 CTCAGTGTGCTTT 7108
Db 1011 CTCAGTGTGCTTT 1024
```

```
RESULT 14
US-08-036-210-8
; Sequence 8, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/036,210
FILING DATE: 23-MAR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cdna
US-08-036-210-8

Query Match 7.5%; Score 532; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.6e-132;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 6437 ACACAAAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTAATGGAGGATGTTTC 6496
DB 1 ACACAAAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTAATGGAGGATGTTTC 60
QY 6497 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGATGGGATTCGATGACTGTTTC 6556
DB 61 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGATGGGATTCGATGACTGTTTC 120
QY 6557 GACAGTGTAACCTTTACTGCTGGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCTAA 6616
DB 121 GACAGTGTAACCTTTACTGCTGGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCTAA 180
QY 6617 TTCACTTTGTGAAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 6676
DB 181 TTCACTTTGTGAAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 240
QY 6677 ACTGAGTGCCTGGAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 6736
DB 241 ACTGAGTGCCTGGAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 300
QY 6737 ATATAAATGACCAATGATTTTGGATATATATGAGTAACTGTTTAACTTCACTTCAAGTCA 6796
DB 301 ATATAAATGACCAATGATTTTGGATATATATGAGTAACTGTTTAACTTCACTTCAAGTCA 420
QY 6857 TCTTATCAATAAGGAGTAACTGAGCCCATCTGTTTGTAACTTCACTTCAAGTCA 6916
DB 421 TCTTATCAATAAGGAGTAACTGAGCCCATCTGTTTGTAACTTCACTTCAAGTCA 480
QY 6917 AGATGGACTCTTTGGAGCCCATGGAAGTGTGAGCTTGAATGGGAAGA 6968
DB 481 AGATGGACTCTTTGGAGCCCATGGAAGTGTGAGCTTGAATGGGAAGA 532
```

RESULT 15
US-08-449-609-8
Sequence 8, Application US/08449609
Patent No. 5952212
GENERAL INFORMATION:
APPLICANT: Moller, Niels P.H.
APPLICANT: Moller, Karin B.
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
TITLE OF INVENTION: PHOSPHATASE

NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSES: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,609
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/036,210
APPLICATION NUMBER: 23-MAR-1993
FILING DATE: 23-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cdna
US-08-449-609-8

Query Match 7.5%; Score 532; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.6e-132;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 6437 ACACAAAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTAATGGAGGATGTTTC 6496
DB 1 ACACAAAGCCAGTACTGCTTTGGAGATATAGTGAATCAAGCTAATGGAGGATGTTTC 60
QY 6497 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGATGGGATTCGATGACTGTTTC 6556
DB 61 AAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGATGGGATTCGATGACTGTTTC 120
QY 6557 GACAGTGTAACCTTTACTGCTGGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCTAA 6616
DB 121 GACAGTGTAACCTTTACTGCTGGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCTAA 180
QY 6617 TTCACTTTGTGAAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 6676
DB 181 TTCACTTTGTGAAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 240
QY 6677 ACTGAGTGCCTGGAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 6736
DB 241 ACTGAGTGCCTGGAGTTGGTTCGAGCAAGCGGACATGACACCATGATGATGTTTC 300
QY 6737 ATATAAATGACCAATGATTTTGGATATATATGAGTAACTGTTTAACTTCACTTCAAGTCA 6796
DB 301 ATATAAATGACCAATGATTTTGGATATATATGAGTAACTGTTTAACTTCACTTCAAGTCA 360
QY 6797 GAATGTGATGTCAGAAATCTGGCAGATATATCTTTTACACAGTGCATTTCTGGATC 6856
DB 361 GAATGTGATGTCAGAAATCTGGCAGATATATCTTTTACACAGTGCATTTCTGGATC 420
QY 6857 TCTTATCAATAAGGAGTAACTGAGCCCATCTGTTTGTAACTTCACTTCAAGTCA 6916
DB 421 TCTTATCAATAAGGAGTAACTGAGCCCATCTGTTTGTAACTTCACTTCAAGTCA 480
```

Qy 6917 AGATGGACTCTTTGGAGCCCATCGAAGGTGATGCTTGAGCTTGAATGGGAAGA 6968
Db 481 AGATGGACTCTTTGGAGCCCATCGAAGGTGATGCTTGAGCTTGAATGGGAAGA 532

Search completed: July 2, 2005, 21:39:34
Job time : 1621 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 12:07:45 ; Search time 5469 Seconds
(without alignments)
8151.890 Million cell updates/sec

Title: US-10-673-885-1
Perfect score: 7108
Sequence: 1 taattgtacttgcagaa.....gcttctcactgtgccttt 7108

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues
Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
23: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7108	100.0	7108	10	US-09-822-871-1
2	7108	100.0	7108	18	US-10-673-885-1
3	6799.8	95.7	7225	18	US-10-466-759-10
4	6683.6	94.0	6903	17	US-10-087-684-9
5	6683.6	94.0	6903	17	US-10-218-779-9
6	6453	90.8	6994	17	US-10-087-684-5
7	6453	90.8	6994	17	US-10-218-779-5

8	2618.6	36.8	3973	15	US-10-314-232-21	Sequence 21, Appl
9	2551	35.9	2565	17	US-10-087-684-7	Sequence 7, Appl
10	2551	35.9	2565	17	US-10-218-779-7	Sequence 7, Appl
11	1118.8	15.7	2692	15	US-10-314-232-14	Sequence 14, Appl
12	1071.6	15.1	2309	15	US-10-314-232-10	Sequence 10, Appl
13	789.6	11.1	2173	15	US-10-314-232-6	Sequence 6, Appl
14	532	7.5	532	15	US-10-314-232-8	Sequence 8, Appl
15	319.4	4.5	321	15	US-10-314-232-3	Sequence 3, Appl
16	315	4.4	254366	10	US-09-822-871-3	Sequence 3, Appl
17	315	4.4	254366	18	US-10-673-885-3	Sequence 3, Appl
18	193.8	2.7	6329	21	US-10-497-692-3	Sequence 3, Appl
19	192.2	2.7	936	18	US-10-634-027-5	Sequence 5, Appl
20	192.2	2.7	1011	18	US-10-634-027-3	Sequence 3, Appl
21	192.2	2.7	5069	17	US-10-331-496A-74	Sequence 74, Appl
22	192.2	2.7	5415	19	US-10-775-169-153	Sequence 153, Appl
23	192.2	2.7	5415	19	US-10-789-241-33	Sequence 33, Appl
24	192.2	2.7	6075	18	US-10-641-643-1366	Sequence 1366, Ap
25	192.2	2.7	6075	18	US-10-634-027-1	Sequence 1, Appl
26	192.2	2.7	6331	17	US-10-159-563-238	Sequence 238, Appl
27	192.2	2.7	7444	10	US-09-909-567B-37	Sequence 37, Appl
28	192.2	2.7	7444	16	US-10-252-157-477	Sequence 477, Appl
29	192.2	2.7	7879	14	US-10-198-846-9851	Sequence 9851, Ap
30	180.8	2.5	4871	21	US-10-486-706-242	Sequence 242, Appl
31	176.2	2.5	6314	17	US-10-366-547-48	Sequence 48, Appl
32	171.8	2.4	4756	17	US-10-366-547-43	Sequence 43, Appl
33	171.8	2.4	4756	18	US-10-641-643-1461	Sequence 1461, Ap
34	171.8	2.4	5117	15	US-10-390-501-1	Sequence 1, Appl
35	171.8	2.4	5117	17	US-10-366-547-41	Sequence 41, Appl
36	171.8	2.4	5117	19	US-10-723-606-1	Sequence 1, Appl
37	169.2	2.4	1752	15	US-10-005-220-1	Sequence 1, Appl
38	167	2.3	4078	17	US-10-305-720-1134	Sequence 1134, Ap
39	167	2.3	4323	11	US-09-997-722-48	Sequence 48, Appl
40	167	2.3	4651	9	US-09-887-669-4	Sequence 4, Appl
41	167	2.3	4651	18	US-10-641-643-945	Sequence 945, Appl
42	167	2.3	5775	11	US-09-997-722-47	Sequence 47, Appl
43	167	2.3	5775	17	US-10-189-429-4	Sequence 4, Appl
44	165.4	2.3	4426	17	US-10-366-547-46	Sequence 46, Appl
45	156.4	2.2	6115	20	US-10-723-860-5538	Sequence 5538, Ap

ALIGNMENTS

RESULT 1
US-09-822-871-1
; Sequence 1, Application US/09822871
; Publication No. US20030099942A1
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO01219
; CURRENT APPLICATION NUMBER: US/09/822,871
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 7108
; TYPE: DNA
; ORGANISM: Human
US-09-822-871-1

Query Match 100.0%; Score 7108; DB 10; Length 7108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TAATTGTGCTACTTGCAGAGATCTGCTTTAAATCAATTAATGCGAGCAACATTTCTCT	60
DB	1	TAATTGTGCTACTTGCAGAGATCTGCTTTAAATCAATTAATGCGAGCAACATTTCTCT	60
QY	61	CTAGAGCCATCAATGTGATTTCTACTGGCTGGAATAAATGTAATAAGATGATTTCTTATC	120

Db 61 CTAGAGCCATCAATGTGATTTCTACTGGCTCAAAAAATGTAAATAAAGATGGAATTTTCTTATC 120
Qy 121 ATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTGATGTTTCCAAATGTCGTT 180
Db 121 ATTTTCTTTTACTTTTATTTGGGACTTCAGAGACACAGGTGATGTTTCCAAATGTCGTT 180
Qy 181 CTTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAAACATACACCTCACCTGTT 240
Db 181 CTTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAAACATACACCTCACCTGTT 240
Qy 241 ACTAGAATAGTGACCAAAATGTAAACAAACAGGGCCCTCCAGTCTTCTTAGCCGGGAA 300
Db 241 ACTAGAATAGTGACCAAAATGTAAACAAACAGGGCCCTCCAGTCTTCTTAGCCGGGAA 300
Qy 301 AGAGTCGGATCTGCTGGGATCTTCTGTCTCGAATACACACCTAAATCCAAATGGAAGG 360
Db 301 AGAGTCGGATCTGCTGGGATCTTCTGTCTCGAATACACACCTAAATCCAAATGGAAGG 360
Qy 361 ATTATATCTTACATTTGTCACAAATATAAGGAAGTTTGTCCGTGGATGCAACAGTATATACA 420
Db 361 ATTATATCTTACATTTGTCACAAATATAAGGAAGTTTGTCCGTGGATGCAACAGTATATACA 420
Qy 421 CAACTCAGATCAAGCCAGACAGTCTGGAAGTTCTTTTCTTAATCTTAATCTTGAACA 480
Db 421 CAACTCAGATCAAGCCAGACAGTCTGGAAGTTCTTTTCTTAATCTTAATCTTGAACA 480
Qy 481 ACATATGAATTAAGGTTCTGCTGAAACAGTCTGGCAATGGAGTGTATGATCCA 540
Db 481 ACATATGAATTAAGGTTCTGCTGAAACAGTCTGGCAATGGAGTGTATGATCCA 540
Qy 541 TTTCTCTTCCAAACTCGAAGAGTGTCCAGGAAAGTGGTGAATCTCCAGTCTGAGGCC 600
Db 541 TTTCTCTTCCAAACTCGAAGAGTGTCCAGGAAAGTGGTGAATCTCCAGTCTGAGGCC 600
Qy 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGCAACCAATGGCAAAAT 660
Db 601 TACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGCAACCAATGGCAAAAT 660
Qy 661 ACCAGCTTCAAGTATAGTGTCAACATGCCAGAGTGGATAGTAGTGAAGATGTCTCA 720
Db 661 ACCAGCTTCAAGTATAGTGTCAACATGCCAGAGTGGATAGTAGTGAAGATGTCTCA 720
Qy 721 ATCAGAGTAGAGACATTTTGACTGGGAAATGGCCAGAGTCAATGAGATAGTGAATCT 780
Db 721 ATCAGAGTAGAGACATTTTGACTGGGAAATGGCCAGAGTCAATGAGATAGTGAATCT 780
Qy 781 TTTTATAGGAGTACAGCCAGCCCTTCTCCAACTTTGGTGAAGTTACACCTCCATCGGCT 840
Db 781 TTTTATAGGAGTACAGCCAGCCCTTCTCCAACTTTGGTGAAGTTACACCTCCATCGGCT 840
Qy 841 ACCACATTCATCAAGCAGCTTGACACAGATGAGATCAGCTCTGTGTGGAAGAGCT 900
Db 841 ACCACATTCATCAAGCAGCTTGACACAGATGAGATCAGCTCTGTGTGGAAGAGCT 900
Qy 901 ATCAGTTTGTAGTGACACATTTGAGACCTTATACAAATATCTTTTGAAGTTTCAGCT 960
Db 901 ATCAGTTTGTAGTGACACATTTGAGACCTTATACAAATATCTTTTGAAGTTTCAGCT 960
Qy 961 GTTACAACCTGAAGCAGTTTATTTGATAGTACGATTTGTCCAGAACCCAGAAATCAGTGCCT 1020
Db 961 GTTACAACCTGAAGCAGTTTATTTGATAGTACGATTTGTCCAGAACCCAGAAATCAGTGCCT 1020
Qy 1021 GAAGNACCAACCAAACTGCGTACAGGCAACATCAAGGAAAGTCCCTTTCAATTTTA 1080
Db 1021 GAAGNACCAACCAAACTGCGTACAGGCAACATCAAGGAAAGTCCCTTTCAATTTTA 1080
Qy 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAATTAATGGAACA 1140
Db 1081 TGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAATTAATGGAACA 1140
Qy 1141 TCAGTTCGATTTTGGATTAACAGCAACAAAGACCTCAAGTTTGCATTTCACTAACCTAACA 1200
Db 1141 TCAGTTCGATTTTGGATTAACAGCAACAAAGACCTCAAGTTTGCATTTCACTAACCTAACA 1200

Qy 1201 CCATTTTCAATCTATGATGTCTATATTGCGGCTGAAACCCAGTGCAGGAGCTGGGCCAAG 1260
Db 1201 CCATTTTCAATCTATGATGTCTATATTGCGGCTGAAACCCAGTGCAGGAGCTGGGCCAAG 1260
Qy 1261 TCAAAATATTTTCAGTATTTCACTCCACCAGATGTTCCAGGGGAGTGTGTTGATTTACAACCT 1320
Db 1261 TCAAAATATTTTCAGTATTTCACTCCACCAGATGTTCCAGGGGAGTGTGTTGATTTACAACCT 1320
Qy 1321 GCAGAGGTAGAAATCCACGCAAGTAAGAATTTCTTGAAGAAACCAAGCAACCAAAATGA 1380
Db 1321 GCAGAGGTAGAAATCCACGCAAGTAAGAATTTCTTGAAGAAACCAAGCAACCAAAATGA 1380
Qy 1381 ATTATTAACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAAAT 1440
Db 1381 ATTATTAACCAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTTGGAAAT 1440
Qy 1441 ACTTTGCTCCTCTGGAATTAATAGATATATAAATGACCCCATGGCTCCAGAAATTTGGAAC 1500
Db 1441 ACTTTGCTCCTCTGGAATTAATAGATATATAAATGACCCCATGGCTCCAGAAATTTGGAAC 1500
Qy 1501 ATAGTAGAGCCAAATGGTAGGATTAATGAGGGTTCAGCAGAGATGTGCTGACCTTCAC 1560
Db 1501 ATAGTAGAGCCAAATGGTAGGATTAATGAGGGTTCAGCAGAGATGTGCTGACCTTCAC 1560
Qy 1561 TCACCTGCTACATTTTATATAACAGCCATCCAGATAAAAACTTTCTGCAAGGAATAGA 1620
Db 1561 TCACCTGCTACATTTTATATAACAGCCATCCAGATAAAAACTTTCTGCAAGGAATAGA 1620
Qy 1621 GCTGAAGACCCAGACTTCACAGTGTGAATCTACAAGGAATCAGTATATTTACTGACATTCGA 1680
Db 1621 GCTGAAGACCCAGACTTCACAGTGTGAATCTACAAGGAATCAGTATATTTACTGACATTCGA 1680
Qy 1681 GCTGAACAGCTGTCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCAGATGATG 1740
Db 1681 GCTGAACAGCTGTCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCAGATGATG 1740
Qy 1741 GTATCTGCTTTTACCACATCATGGGAGAGGACCAACAGTTCTCAGTGTGAGGACAGT 1800
Db 1741 GTATCTGCTTTTACCACATCATGGGAGAGGACCAACAGTTCTCAGTGTGAGGACAGT 1800
Qy 1801 CAGCAAGTGCAGCTCCATTTAAAAATTATAAACTATAAAAAATTTAGTTCTTCATCTATT 1860
Db 1801 CAGCAAGTGCAGCTCCATTTAAAAATTATAAACTATAAAAAATTTAGTTCTTCATCTATT 1860
Qy 1861 TTGTTATATTGGGATCTCCAGATATCCCAATGGAATAAATTAATCCTACTATACGATTTAT 1920
Db 1861 TTGTTATATTGGGATCTCCAGATATCCCAATGGAATAAATTAATCCTACTATACGATTTAT 1920
Qy 1921 GCAATGGAATTTGGATACAAACAGACTTCCAGATAAATACCATAAGATAAAGCTTTCTC 1980
Db 1921 GCAATGGAATTTGGATACAAACAGACTTCCAGATAAATACCATAAGATAAAGCTTTCTC 1980
Qy 1981 ATAAACAGGGTTAAAGAAATACAAAATACAAAATGAGAGTGGCAGCTCAACCCACGAT 2040
Db 1981 ATAAACAGGGTTAAAGAAATACAAAATACAAAATGAGAGTGGCAGCTCAACCCACGAT 2040
Qy 2041 GGAGAAAGTTCTTTCTGGAAGAAATGACATCTTTTGTGAGAACTTCAGAGATGAACCG 2100
Db 2041 GGAGAAAGTTCTTTCTGGAAGAAATGACATCTTTTGTGAGAACTTCAGAGATGAACCG 2100
Qy 2101 GAATCATCACTCAAGATGTCGAAGTAATGATGTTTACCGCAGATGAATAAGGTTGAAG 2160
Db 2101 GAATCATCACTCAAGATGTCGAAGTAATGATGTTTACCGCAGATGAATAAGGTTGAAG 2160
Qy 2161 TGGTCAACCCCGGAAAGCCCAATGCGGATCAITTTTGTCTTATGAAGTGTCTATATAAAT 2220
Db 2161 TGGTCAACCCCGGAAAGCCCAATGCGGATCAITTTTGTCTTATGAAGTGTCTATATAAAT 2220
Qy 2221 ATAGATATCTTTTATATATGAAGAACACATCAACAGACATATATTAAGGAACTTAAGA 2280
Db 2221 ATAGATATCTTTTATATATGAAGAACACATCAACAGACATATATTAAGGAACTTAAGA 2280

QY 2281 CCTCACACCTCTATAACATTTCTGTAAAGTCTTACACAGATTGGTCATGCGCAATCAG 2340
DB CCTCACACCTCTATAACATTTCTGTAAAGTCTTACACAGATTGGTCATGCGCAATCAG 2340
QY 2341 GTATCTTTCTTACTCTCTGTAAAGTCTCGGAGCTGTGCTGATAGTACACAGAAAT 2400
DB GTATCTTTCTTACTCTCTGTAAAGTCTCGGAGCTGTGCTGATAGTACACAGAAAT 2400
QY 2401 ATCACTTACAAAATATTTCTTCTGGAGAGTTGAGCTATCATTTCTCCCCCAAGTAGT 2460
DB ATCACTTACAAAATATTTCTTCTGGAGAGTTGAGCTATCATTTCTCCCCCAAGTAGT 2460
QY 2461 CCCAATGGATCAATAAAAAATATACAATTTATCTCAAGAGAGTAATGGAATGAGGAA 2520
DB CCCAATGGATCAATAAAAAATATACAATTTATCTCARGAGAGTAATGGAATGAGGAA 2520
QY 2521 AGAATCTATAAATACAACTCTTTTAAACCAAAATTTAAAGTACTGAAAGAAATATACCCAA 2580
DB AGAATCTATAAATACAACTCTTTTAAACCAAAATTTAAAGTACTGAAAGAAATATACCCAA 2580
QY 2581 TATATCAITGAGGTGTCTGTAGTACACTGAAAGGTGAAGGAGTTCCGAGTGTCTCCATA 2640
DB TATATCAITGAGGTGTCTGTAGTACACTGAAAGGTGAAGGAGTTCCGAGTGTCTCCATA 2640
QY 2641 AGTATCTGACGGAGAGAGTCTCTGTATTTCTCCCTCAAGACTTCTCTGTAAACAG 2700
DB AGTATCTGACGGAGAGAGTCTCTGTATTTCTCCCTCAAGACTTCTCTGTAAACAG 2700
QY 2701 TTGCTGGTGTACGGTGAAGTGTCTATGCGCAACACCTCGAGCCAAATGGAATATC 2760
DB TTGCTGGTGTACGGTGAAGTGTCTATGCGCAACACCTCGAGCCAAATGGAATATC 2760
QY 2761 CTTTATTACACAGTTTATGTCTGGAATAGATCATTAATAAATCTTAAATGTCACTGAA 2820
DB CTTTATTACACAGTTTATGTCTGGAATAGATCATTAATAAATCTTAAATGTCACTGAA 2820
QY 2821 ACATCAITGGAGTTATCAAGTTTGAATTAATGTTGAATACAGTCTTATGTAACAGCT 2880
DB ACATCAITGGAGTTATCAAGTTTGAATTAATGTTGAATACAGTCTTATGTAACAGCT 2880
QY 2881 AGCACCAGATTTGGTATGGGAAACAGGAGCAATATCATTTAGCTTCCAAACACAGAG 2940
DB AGCACCAGATTTGGTATGGGAAACAGGAGCAATATCATTTAGCTTCCAAACACAGAG 2940
QY 2941 GGAGCAACCAAGGATCTCTCCAAAGATGTTTATATGCAAACTCTAGTTCTTCAATCAATA 3000
DB GGAGCAACCAAGGATCTCTCCAAAGATGTTTATATGCAAACTCTAGTTCTTCAATCAATA 3000
QY 3001 ATTTCTTTCTGGACACTCTCTCAAAACCTTAATGGGATTAACAATTTACTCTGTTTAT 3060
DB ATTTCTTTCTGGACACTCTCTCAAAACCTTAATGGGATTAACAATTTACTCTGTTTAT 3060
QY 3061 TACAGAAATACCTCAGTACTTTTATGCAAGATTTTACACTCCATGAACTTAACTCAATGAC 3120
DB TACAGAAATACCTCAGTACTTTTATGCAAGATTTTACACTCCATGAACTTAACTCAATGAC 3120
QY 3121 TTTGCAATATGACTGTATCCCAATATATAGATAAACTGACAAATTTAGCTACTATACA 3180
DB TTTGCAATATGACTGTATCCCAATATATAGATAAACTGACAAATTTAGCTACTATACA 3180
QY 3181 TTTTGGTTAACGCAAGTACTCTTCAAGTTGGAATGGAATTAAGGAGTACATCAATTGAA 3240
DB TTTTGGTTAACGCAAGTACTCTTCAAGTTGGAATGGAATTAAGGAGTACATCAATTGAA 3240
QY 3241 GTATACACAGATCAAGACATACCTGAGGTTTGTGGAAACCTGACTTACGAATCCATT 3300
DB GTATACACAGATCAAGACATACCTGAGGTTTGTGGAAACCTGACTTACGAATCCATT 3300
QY 3301 TGTCAACTGCAATAAATGTAAGTCTGGTCCCAACCGGCTCAACAAACGCTCTAGTCTTC 3360
DB TGTCAACTGCAATAAATGTAAGTCTGGTCCCAACCGGCTCAACAAACGCTCTAGTCTTC 3360
QY 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCTCGCCATGTGAGACCACCTCTTGTTACA 3420

DB 3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCTCGCCATGTGAGACCACCTCTTGTTACA 3420
QY 3421 TATGAGAGAGCATATATTTTGTATATCTGGAATAATCACTGATATATATATAAATAAT 3480
DB TATGAGAGAGCATATATTTTGTATATCTGGAATAATCACTGATATATATATAAATAAT 3480
QY 3481 ACTCATCAACAGAAAGGATTTCTGTATACCTATFACCTGCGCAGCTATACATCAAGACT 3540
DB ACTCATCAACAGAAAGGATTTCTGTATACCTATFACCTGCGCAGCTATACATCAAGACT 3540
QY 3541 GAAGAGATGTCCCAAGAACTTCAACCAATTAATCAACACTTTTAAACACCTTTCTCTACC 3600
DB GAAGAGATGTCCCAAGAACTTCAACCAATTAATCAACACTTTTAAACACCTTTCTCTACC 3600
QY 3601 TCAGTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGCATATATAAGTTATGAT 3660
DB TCAGTCTCTTATCATGGGATCCCCAGTAAAGCCAAATGGTGCATATATAAGTTATGAT 3660
QY 3661 TTAACCTTTACAGAGCAAAATGAAATATTTCTTTTACACAGATGAGTGCCTGAGTACACT 3720
DB TTAACCTTTACAGAGCAAAATGAAATATTTCTTTTACACAGATGAGTGCCTGAGTACACT 3720
QY 3721 TTGGAAGAGCTTTCAACATTTACATATATAGCTTTTGTGCTGCCGCAAGACTAGAAA 3780
DB TTGGAAGAGCTTTCAACATTTACATATATAGCTTTTGTGCTGCCGCAAGACTAGAAA 3780
QY 3781 GGACTTGGTCTTCCAGTATTTCTTTTACACAGATGAGTGCCTGAGTACACT 3840
DB GGACTTGGTCTTCCAGTATTTCTTTTACACAGATGAGTGCCTGAGTACACT 3840
QY 3841 CCACAAAATTTGAATTAATCAACTGTACTTCCAGCTTTTGTATGCTGAAATGGAGCCCA 3900
DB CCACAAAATTTGAATTAATCAACTGTACTTCCAGCTTTTGTATGCTGAAATGGAGCCCA 3900
QY 3901 AGTCTCTTCCAGTGGTATTTGTTAAAGTATATAGTTTAAATTCATGAACTGAACT 3960
DB AGTCTCTTCCAGTGGTATTTGTTAAAGTATATAGTTTAAATTCATGAACTGAACT 3960
QY 3961 GACACTATATATTAAGAAATATATCAGGATTTAAATCAAGCCAACTCTGTTGAGCTG 4020
DB GACACTATATATTAAGAAATATATCAGGATTTAAATCAAGCCAACTCTGTTGAGCTG 4020
QY 4021 GAACAGTCAAGCACTCTATCTGCTGTATCTGCTTACCAAAAGTTGGAATGGCAAT 4080
DB GAACAGTCAAGCACTCTATCTGCTGTATCTGCTTACCAAAAGTTGGAATGGCAAT 4080
QY 4081 CAAATTTAGTATGTAAGTAAATTTCAACCCCAAGAAATCAGTTCCAGATGTCTGCAAGAT 4140
DB CAAATTTAGTATGTAAGTAAATTTCAACCCCAAGAAATCAGTTCCAGATGTCTGCAAGAT 4140
QY 4141 ATGCAAGTCAAGCACTGAGTGGCAGTCAAGTTTAAATGGATGGATCCACCAAAAG 4200
DB ATGCAAGTCAAGCACTGAGTGGCAGTCAAGTTTAAATGGATGGATCCACCAAAAG 4200
QY 4201 GCAATTTGAATTAATCAAGCAGTATATGGTAAACAGTGTGGAAGAAATTTCTA 4260
DB GCAATTTGAATTAATCAAGCAGTATATGGTAAACAGTGTGGAAGAAATTTCTA 4260
QY 4261 CCCCAAGATCAATGTACACTTTTCAAAAGCTTTTGGCAATACCTCATATGTCTTTAAA 4320
DB CCCCAAGATCAATGTACACTTTTCAAAAGCTTTTGGCAATACCTCATATGTCTTTAAA 4320
QY 4321 GTAAGAGCTTCAACCTCAGTGTGATGAAGCAATGCAATGCAATGTCAGCACATA 4380
DB GTAAGAGCTTCAACCTCAGTGTGATGAAGCAATGCAATGCAATGTCAGCACATA 4380
QY 4381 CCTGAAACAGTTTCCAGTGTCTCCCAAAATATTTGCTTTTCTGATGTTTCTGAGTCAACTAGT 4440
DB CCTGAAACAGTTTCCAGTGTCTCCCAAAATATTTGCTTTTCTGATGTTTCTGAGTCAACTAGT 4440
QY 4441 GCAATTTGACATGGAATGAAGCTTCACTATCTTGGCTACTTTTCAAAATTAACAAATTT 4500
DB GCAATTTGACATGGAATGAAGCTTCACTATCTTGGCTACTTTTCAAAATTAACAAATTT 4500

Db 4441 GCAACATTGACATGGATGAAGCCTGACACTATCTCTTGGCTACTTTTCAAAAATTACAAAATT 4500
QY 4501 ACCACTCAACTTCGTCTCAAAAATGCAAGAAATGGAATCCGAAGAAATGCTGTGAATAT 4560
Db 4501 ACCACTCAACTTCGTCTCAAAAATGCAAGAAATGGAATCCGAAGAAATGCTGTGAATAT 4560
QY 4561 CAAAAAATTCAATPACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGATGATTAAG 4620
Db 4561 CAAAAAATTCAATPACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGATGATTAAG 4620
QY 4621 AAATTTAGATGGATATAGATTCGAAGTGGCTGCGCAGCACCAATGCTGGCTATGCGCAATGCT 4680
Db 4621 AAATTTAGATGGATATAGATTCGAAGTGGCTGCGCAGCACCAATGCTGGCTATGCGCAATGCT 4680
QY 4681 TCAAACTGATTTCTACAAAACCTCTGCTGGCCCTCCAGATGGCTCTCTGAAAATGTT 4740
Db 4681 TCAAACTGATTTCTACAAAACCTCTGCTGGCCCTCCAGATGGCTCTCTGAAAATGTT 4740
QY 4741 CATGTAGTAGCAACATCACTTTTAGCATCAGCATAAAGCTGGAGTGAACCTGCTGTCAAT 4800
Db 4741 CATGTAGTAGCAACATCACTTTTAGCATCAGCATAAAGCTGGAGTGAACCTGCTGTCAAT 4800
QY 4801 ACTGGACCAACATGTTATCTGATTTGATGTCAAATCGGTAGATAATGATGAATTTAATAATA 4860
Db 4801 ACTGGACCAACATGTTATCTGATTTGATGTCAAATCGGTAGATAATGATGAATTTAATAATA 4860
QY 4861 TCCCTCATCAAGTCAAAATGAAGAAATAAACCATAGATAATTAAGATTTAGAAATATTC 4920
Db 4861 TCCCTCATCAAGTCAAAATGAAGAAATAAACCATAGATAATTAAGATTTAGAAATATTC 4920
QY 4921 ACAAGGATTTCTGTAGTGATCACTGCAATTTACTGGGAACATTTAGTGTCTCATATGTAGAA 4980
Db 4921 ACAAGGATTTCTGTAGTGATCACTGCAATTTACTGGGAACATTTAGTGTCTCATATGTAGAA 4980
QY 4981 GGGAGTCAAGTGTGAAATGATTTGTTACTTACTTTAGAAATCAGCCCAAGGACCCACCT 5040
Db 4981 GGGAGTCAAGTGTGAAATGATTTGTTACTTACTTTAGAAATCAGCCCAAGGACCCACCT 5040
QY 5041 AACACATGACATTTTCAGAAATACCCAGATGAAGTTACAAAATTTCAATTAACGTTCTCT 5100
Db 5041 AACACATGACATTTTCAGAAATACCCAGATGAAGTTACAAAATTTCAATTAACGTTCTCT 5100
QY 5101 CTTCTCTTCAACTAATGGAATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGAT 5160
Db 5101 CTTCTCTTCAACTAATGGAATATCCAAAGTATATCAAGCTCTGGTTTACCGAGAAGAT 5160
QY 5161 GATCTTACTGCTGTCAGATTCACACCTCAGTATTTATACAGAAAACCAACACATTCGTC 5220
Db 5161 GATCTTACTGCTGTCAGATTCACACCTCAGTATTTATACAGAAAACCAACACATTCGTC 5220
QY 5221 ATTGCAATGCTAGAGGACTAAAGGTGGACATACATACATATCAGTGTGTTACCGCAGTC 5280
Db 5221 ATTGCAATGCTAGAGGACTAAAGGTGGACATACATACATATCAGTGTGTTACCGCAGTC 5280
QY 5281 AATAGTGTGTCAGTCCAAAGGTTCCGATGAGAAATACCATGATATCAAGCTCCA 5340
Db 5281 AATAGTGTGTCAGTCCAAAGGTTCCGATGAGAAATACCATGATATCAAGCTCCA 5340
QY 5341 GCACGACCAAAACCAAAACCAACCCCTATTTATGATGCCACAGGAAAACCTGCTGTGACT 5400
Db 5341 GCACGACCAAAACCAAAACCAACCCCTATTTATGATGCCACAGGAAAACCTGCTGTGACT 5400
QY 5401 TCAACAAATTTACATATCAGATTCGCAATATGTTTACTACAGTGTGATCATGAGCCAAATA 5460
Db 5401 TCAACAAATTTACATATCAGATTCGCAATATGTTTACTACAGTGTGATCATGAGCCAAATA 5460
QY 5461 AAAAAATGTAAGTGTCTCGACAGAAACAGAGCTTCAGCATGATGGAATGTAAACAAAG 5520
Db 5461 AAAAAATGTAAGTGTCTCGACAGAAACAGAGCTTCAGCATGATGGAATGTAAACAAAG 5520
QY 5521 TGGTATGATGCAATTTTAAATGAAGCAAGGCCATATTTTCAAAATGAAGGCTTTCTTAC 5580
Db 5521 TGGTATGATGCAATTTTAAATGAAGCAAGGCCATATTTTCAAAATGAAGGCTTTCTTAC 5580

QY 5581 CCTCCATGTACAGAAAGAACAAAAGTTTAGTGGCAATGAAGAAATCTTACATCATAGT 5640
Db 5581 CCTCCATGTACAGAAAGAACAAAAGTTTAGTGGCAATGAAGAAATCTTACATCATAGT 5640
QY 5641 GCTGATAATGATGATGATTCCTGGCAATGAAGAATAAAATTTGCAATGGACACATGAAA 5700
Db 5641 GCTGATAATGATGATGATTCCTGGCAATGAAGAATAAAATTTGCAATGGACACATGAAA 5700
QY 5701 CAAAAAAGCAATACCTTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTACTGAC 5760
Db 5701 CAAAAAAGCAATACCTTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTACTGAC 5760
QY 5761 TCTGATATTTCTGACCTCTTAAAGCTTTTAGGGGAAGGACTTTTCAGAAAAGAACCTGTAGAG 5820
Db 5761 TCTGATATTTCTGACCTCTTAAAGCTTTTAGGGGAAGGACTTTTCAGAAAAGAACCTGTAGAG 5820
QY 5821 ATCATCTTTTCCGTCACTTTTGTGTATCTTTCAATAATTTCTCTTGGAAACAGCTATTTTT 5880
Db 5821 ATCATCTTTTCCGTCACTTTTGTGTATCTTTCAATAATTTCTCTTGGAAACAGCTATTTTT 5880
QY 5881 GCATTTGCAAGAAATTCGACAGAGCAAGAAAGAGTGGCACATATCTCTCTCAGATGCA 5940
Db 5881 GCATTTGCAAGAAATTCGACAGAGCAAGAAAGAGTGGCACATATCTCTCTCAGATGCA 5940
QY 5941 GAAATTTATTTGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACCTG 6000
Db 5941 GAAATTTATTTGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACCTG 6000
QY 6001 AAGGACGAGAGATTAACCGGCCAATAAGCAAGAAATCTTCTCTGCAACATGTTGAAGAG 6060
Db 6001 AAGGACGAGAGATTAACCGGCCAATAAGCAAGAAATCTTCTCTGCAACATGTTGAAGAG 6060
QY 6061 CTTTGCACAAACAAACCTAAAGTTTCAAGAAATTTTCGGAATTTACCAAAATTTCTT 6120
Db 6061 CTTTGCACAAACAAACCTAAAGTTTCAAGAAATTTTCGGAATTTACCAAAATTTCTT 6120
QY 6121 CAGGATCTTCTTCAACTGATCTGCTCTGCTGCAATAGAGCAAAACCCGTTTCCCA 6180
Db 6121 CAGGATCTTCTTCAACTGATCTGCTCTGCTGCAATAGAGCAAAACCCGTTTCCCA 6180
QY 6181 AACATAAACCAATATAATAATAAACAGAGTAAAGCTGATAGCTGACGCTAGTGTCCA 6240
Db 6181 AACATAAACCAATATAATAATAAACAGAGTAAAGCTGATAGCTGACGCTAGTGTCCA 6240
QY 6241 GGTTCGATATATTAATGCGCAGCTATATTTCTGTTATTTATGTTCCAAATGAAATTTAT 6300
Db 6241 GGTTCGATATATTAATGCGCAGCTATATTTCTGTTATTTATGTTCCAAATGAAATTTAT 6300
QY 6301 GCTACTCAAGGTCCTACTACCGAGAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACC 6360
Db 6301 GCTACTCAAGGTCCTACTACCGAGAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACC 6360
QY 6361 AGGGCAAAACCAATTAAGTAACTAAACACAGTGTGTTTGAAGGACGAGTACAGTGCCAT 6420
Db 6361 AGGGCAAAACCAATTAAGTAACTAAACACAGTGTGTTTGAAGGACGAGTACAGTGCCAT 6420
QY 6421 CAGTATTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTTGGAGATATATGATGATTAAG 6480
Db 6421 CAGTATTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTTGGAGATATATGATGATTAAG 6480
QY 6481 CTAATGGAGGATTTCAAAATGATTTGGAATACTACGGATCTGAAATTTGAAAGGATGG 6540
Db 6481 CTAATGGAGGATTTCAAAATGATTTGGAATACTACGGATCTGAAATTTGAAAGGATGG 6540
QY 6541 GATTCATGACTGTTTCGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCTGAG 6600
Db 6541 GATTCATGACTGTTTCGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCTGAG 6600
QY 6601 AACAGGCCCCCTCTAATTTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGGCACAATGACCC 6660
Db 6601 AACAGGCCCCCTCTAATTTCACTTTGTGAAGTTGGTTTCGAGCAAGCAGGGCACAATGACCC 6660

```
QY 6661 ACACCTATGATGTTCTCAGTCTGAGTCTGGAGTGGAGAACTGGAGTCTTTATGCTCTG 6720
Db 6661 ACACCTATGATGTTCTCAGTCTGAGTCTGGAGTGGAGAACTGGAGTCTTTATGCTCTG 6720
QY 6721 GACCATTTAAACACAAATATTAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCT 6780
Db 6721 GACCATTTAAACACAAATATTAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCT 6780
QY 6781 GAACTGAGAAGTGAAGAATGTGCATGGTGCAGAACTGGCACAGTATATCTTTTACAC 6840
Db 6781 GAACTGAGAAGTGAAGAATGTGCATGGTGCAGAACTGGCACAGTATATCTTTTACAC 6840
QY 6841 CAGTGATCTGATCTCTTATCAATTAAGGAAGTAATCAGGCCATCTGTTTGTAAAC 6900
Db 6841 CAGTGATCTGATCTCTTATCAATTAAGGAAGTAATCAGGCCATCTGTTTGTAAAC 6900
QY 6901 TATTGAGCACTTCAGAAAGTGAAGTCTTTGGAGCCATGGAAGTGTGAGCTTGAA 6960
Db 6901 TATTGAGCACTTCAGAAAGTGAAGTCTTTGGAGCCATGGAAGTGTGAGCTTGAA 6960
QY 6961 TGGGAAGAAACCACTATGTAATATTTAGACCAAGGATACAAATGGAAAGATTTTAA 7020
Db 6961 TGGGAAGAAACCACTATGTAATATTTAGACCAAGGATACAAATGGAAAGATTTTAA 7020
QY 7021 ATCCAGGGGCCAAAGTACCCTCATCTTCCGAATTTGAATGTGCAACCTTAAAGAA 7080
Db 7021 ATCCAGGGGCCAAAGTACCCTCATCTTCCGAATTTGAATGTGCAACCTTAAAGAA 7080
QY 7081 ATATCTATGCTCTCTCAGTGGCTTT 7108
Db 7081 ATATCTATGCTCTCTCAGTGGCTTT 7108
```

RESULT 2

```
US-10-673-885-1
; Sequence 1, Application US/10673885
; Publication No. US20040081644A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219DIV
; CURRENT APPLICATION NUMBER: US/10/673,885
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 09/822,871
; PRIOR FILING DATE: 04-02-2001
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 7108
; TYPE: DNA
; ORGANISM: Human
US-10-673-885-1
```

```
Query Match 100.0%; Score 7108; DB 18; Length 7108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAATTGTGTAATTGCGAGGAAGATCTGCTTTAAATCATTAATGCAAGCAACATTTCTCT 60
Db 1 TAATTGTGTAATTGCGAGGAAGATCTGCTTTAAATCATTAATGCAAGCAACATTTCTCT 60
QY 61 CTAGAGCCATCATGATGATCTGCTGGAAGATGTAATGAAGATGATTTCTTATC 120
Db 61 CTAGAGCCATCATGATGATCTGCTGGAAGATGTAATGAAGATGATTTCTTATC 120
QY 121 ATTTTCTTTTACTTTTATTTGGAGCTTCAGAGACACAGGTTGATGTTTCCATGTCGTT 180
Db 121 ATTTTCTTTTACTTTTATTTGGAGCTTCAGAGACACAGGTTGATGTTTCCATGTCGTT 180
QY 181 CCTGGTACTAGTACGATATAACCATCTCTTCAATTTCTTACAAACATACACCTCACTGTT 240
Db 181 CCTGGTACTAGTACGATATAACCATCTCTTCAATTTCTTACAAACATACACCTCACTGTT 240
```

```
Db 181 CCTGGTACTAGTACGATATAACCATCTCTTCAATTTCTTACAAACATACACCTCACTGTT 240
QY 241 ACTAGATAAGTACGACCAAAATGTAACAAACACAGGGCCCTCCAGTCTTCTTAGCCGGGAA 300
Db 241 ACTAGATAAGTACGACCAAAATGTAACAAACACAGGGCCCTCCAGTCTTCTTAGCCGGGAA 300
QY 301 AGAGTCGGATCTGCTGGGATTTCTTCTGCTTGGATACACCACTTAATCAATGGAAGG 360
Db 301 AGAGTCGGATCTGCTGGGATTTCTTCTGCTTGGATACACCACTTAATCAATGGAAGG 360
QY 361 ATTATATCTTCAATTTGTAATATTAAGGAAGTTTGTCCGTGATGCAACAGTATATACA 420
Db 361 ATTATATCTTCAATTTGTAATATTAAGGAAGTTTGTCCGTGATGCAACAGTATATACA 420
QY 421 CAAAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTAAATCTTAAATCTTAAATCT 480
Db 421 CAAAGTCAGATCAAGCCAGACAGTCTGGAAGTTCTTCTTAAATCTTAAATCTTAAATCT 480
QY 481 ACATATGAATTAAGGTTGCTGCTGGAAGTCTGGAAGTCTGGAAGTCTGGAAGTCTGGAAG 540
Db 481 ACATATGAATTAAGGTTGCTGCTGGAAGTCTGGAAGTCTGGAAGTCTGGAAGTCTGGAAG 540
QY 541 TTTCTCTTCCAACTGCGAAGAGTCTCCAGGAAAGTGTGGAATCTCAGTGTGAGGCC 600
Db 541 TTTCTCTTCCAACTGCGAAGAGTGTGGAATCTCAGGAAAGTGTGGAATCTCAGTGTGAGGCC 600
QY 601 TACAAAGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCAACCAATGGCAAAAT 660
Db 601 TACAAAGCTTCAGCAGTTAAGCTGATTTGGTATTTTACCTCGGCAACCAATGGCAAAAT 660
QY 661 ACCAGCTTCAAGATTTAGTGTCAAAATGCGAAGTGGATAGTGTGGAAGATGTCTCA 720
Db 661 ACCAGCTTCAAGATTTAGTGTCAAAATGCGAAGTGGATAGTGTGGAAGATGTCTCA 720
QY 721 ATCAGATAGAGGACATTTTGAATGGGAAATGCGAAGTGGATAGTGTGGAAGATGTCTCA 780
Db 721 ATCAGATAGAGGACATTTTGAATGGGAAATGCGAAGTGGATAGTGTGGAAGATGTCTCA 780
QY 781 TTTTATGAGTACAGCCAGCCCTTCTTCAACCTTGTGAGTGTGAGTGTGAGTGTGAGTGT 840
Db 781 TTTTATGAGTACAGCCAGCCCTTCTTCAACCTTGTGAGTGTGAGTGTGAGTGTGAGTGT 840
QY 841 ACCACATTTCAATGAAGCACTGTGACAGAAATGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
Db 841 ACCACATTTCAATGAAGCACTGTGACAGAAATGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
QY 901 ATCAGTTTGTAGTGTGACACATGAGCCTTATACAAATATCTTTTGAAGTGTGAGTGT 960
Db 901 ATCAGTTTGTAGTGTGACACATGAGCCTTATACAAATATCTTTTGAAGTGTGAGTGT 960
QY 961 GCTCAACCTGAAGCAGGTTTATATGATAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1020
Db 961 GCTCAACCTGAAGCAGGTTTATATGATAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1020
QY 1021 GAAGGACCAACCAAAATGCGTAAACAGCAACATCAGCAAGAAAGTCTTTTCAATTTTA 1080
Db 1021 GAAGGACCAACCAAAATGCGTAAACAGCAACATCAGCAAGAAAGTCTTTTCAATTTTA 1080
QY 1081 TGGGACCCCAACCAATATAGTAAACAGGAAATTTAGTATAGTGTGAAATTTATGAGACCA 1140
Db 1081 TGGGACCCCAACCAATATAGTAAACAGGAAATTTAGTATAGTGTGAAATTTATGAGACCA 1140
QY 1141 TCAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAGTTTGTGATTTCAACCTTAAACA 1200
Db 1141 TCAGGTGCGATTTTGGATAACAGCAACAAAGACCTCAAGTTTGTGATTTCAACCTTAAACA 1200
QY 1201 CCAATTTCAATGATGATGTCTATATTTGGGCTGAAACAGTGTGAGGAGCTGGGCCCAAG 1260
Db 1201 CCAATTTCAATGATGATGTCTATATTTGGGCTGAAACAGTGTGAGGAGCTGGGCCCAAG 1260
QY 1261 TCAATATTTTCAATGATGATGTCTATATTTGGGCTGAAACAGTGTGAGGAGCTGGGCCCAAG 1320
Db 1261 TCAATATTTTCAATGATGATGTCTATATTTGGGCTGAAACAGTGTGAGGAGCTGGGCCCAAG 1320
```


1321 GCAGAGGTAGATCCACGCAAGTAAGAAATTACTTTGGAAGAAACACGACAAACCAATGGA 1380
1321 GCAGAGGTAGATCCACGCAAGTAAGAAATTACTTTGGAAGAAACACGACAAACCAATGGA 1380
1381 ATTATTAAACCAATACCGAGTGAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAT 1440
1381 ATTATTAAACCAATACCGAGTGAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAT 1440
1441 ACTTTGCTCACTGGAAATTAATGAGTATATTAATGACCCCATGGCTCCAGAAATTTGGAAC 1500
1441 ACTTTGCTCACTGGAAATTAATGAGTATATTAATGACCCCATGGCTCCAGAAATTTGGAAC 1500
1501 ATAGTAGAGCAATGTTAGGATTTATAGGGTTTACGACAGAGATGTCGTCTGACCTTCAC 1560
1501 ATAGTAGAGCAATGTTAGGATTTATAGGGTTTACGACAGAGATGTCGTCTGACCTTCAC 1560
1561 TCATTGCTACATTTATATATATAACAGCCATCCAGATATAAAACCTTTCGCAAGGAATAGA 1620
1561 TCATTGCTACATTTATATATATAACAGCCATCCAGATATAAAACCTTTCGCAAGGAATAGA 1620
1621 GCTGAAGACAGACTTCCACGTTGTAACTACAGGAATCAGTATATTAATGACATTGCA 1680
1621 GCTGAAGACAGACTTCCACGTTGTAACTACAGGAATCAGTATATTAATGACATTGCA 1680
1681 GCTGAACAGCTGCTTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACATGATTAGT 1740
1681 GCTGAACAGCTGCTTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACATGATTAGT 1740
1741 GTATCTGCTTTCACCATCATGGAGAGGACCAACCAAGTTCTCAGTGTAGGACACGT 1800
1741 GTATCTGCTTTCACCATCATGGAGAGGACCAACCAAGTTCTCAGTGTAGGACACGT 1800
1801 CAGCAAGTGCAGCTCCATTTAAATTTAACTATATAAAATTAATGATTTCTCATCTATT 1860
1801 CAGCAAGTGCAGCTCCATTTAAATTTAACTATATAAAATTAATGATTTCTCATCTATT 1860
1861 TTGTTATATTTGGATCCTCCAGAAATCCCAATGGAAATAAATCACTATACGATTTAT 1920
1861 TTGTTATATTTGGATCCTCCAGAAATCCCAATGGAAATAAATCACTATACGATTTAT 1920
1921 GCAATGGAATTTGGATCAAAACAGAGCAATCCAGATAACTACCATAGATAACGCTTCTC 1980
1921 GCAATGGAATTTGGATCAAAACAGAGCAATCCAGATAACTACCATAGATAACGCTTCTC 1980
1981 ATAAAGGCTTAAAGAAATACACAAATACAAATGAGAGTGGCAGCCTCAACCCACGAT 2040
1981 ATAAAGGCTTAAAGAAATACACAAATACAAATGAGAGTGGCAGCCTCAACCCACGAT 2040
2041 GGAGAAAGTTCTTTGCTGAAGAAATGACATCTTTGTGAGAACTTCAGAAAGTGAACCG 2100
2041 GGAGAAAGTTCTTTGCTGAAGAAATGACATCTTTGTGAGAACTTCAGAAAGTGAACCG 2100
2101 GAATCATCACTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAATAAGTTGAAG 2160
2101 GAATCATCACTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAATAAGTTGAAG 2160
2161 TGGTCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGAGTGTCTATATAAAAT 2220
2161 TGGTCAACCCGAAAGCCCAATGGGATCATTTATGCTTATGAGTGTCTATATAAAAT 2220
2221 ATAGATACCTTTATATGAGAAACACATCAACAGACATTAATTTAAGGAACCTTAAGA 2280
2221 ATAGATACCTTTATATGAGAAACACATCAACAGACATTAATTTAAGGAACCTTAAGA 2280
2281 CCTCACACCTCTTATAACATTTCTGTAAGGTCTTACACAGATTTTGGTCATGGCAATCAG 2340
2281 CCTCACACCTCTTATAACATTTCTGTAAGGTCTTACACAGATTTTGGTCATGGCAATCAG 2340
2341 GTATCTTTTACTCTCTGTAAGGACTTCGGAGACTGTGCTGATAGTGCCACAGAAAT 2400
2341 GTATCTTTTACTCTCTGTAAGGACTTCGGAGACTGTGCTGATAGTGCCACAGAAAT 2400

2401 ATCACTTCAAAAATATTTCTTCTGGAGAGATTGAGCTATCATCTCTTCCCCCAAGTAGT 2460
2401 ATCACTTCAAAAATATTTCTTCTGGAGAGATTGAGCTATCATCTCTTCCCCCAAGTAGT 2460
2461 CCCAATGGAATCATTAATAAATAATACAATTTATCTCAAGAGAAAGTAATGGAATGAGGAA 2520
2461 CCCAATGGAATCATTAATAAATAATACAATTTATCTCAAGAGAAAGTAATGGAATGAGGAA 2520
2521 AGAATCATTAATAACACCTCTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAA 2580
2521 AGAATCATTAATAACACCTCTTAAACCCAAACATTAAGTACTGAAGAAATATACCCAA 2580
2581 TATATCATTTGAGGTCTCTGCTAGTACACTGAAAGGTGAAGGAGTTGCGAGTCTCCCAT 2640
2581 TATATCATTTGAGGTCTCTGCTAGTACACTGAAAGGTGAAGGAGTTGCGAGTCTCCCAT 2640
2641 AGTATCATGACGAGGAGAGTCTCTGANTCTCCCTCAAGACTTCTCTGTAATAACAG 2700
2641 AGTATCATGACGAGGAGAGTCTCTGANTCTCCCTCAAGACTTCTCTGTAATAACAG 2700
2701 TTGCTCTGCTGTCACGCTGAAGTTGTCATGGCAACCAACCCCTGGAGCCAAATGGAATATC 2760
2701 TTGCTCTGCTGTCACGCTGAAGTTGTCATGGCAACCAACCCCTGGAGCCAAATGGAATATC 2760
2761 CTTTATTTACACAGTTTATGCTCTGGAATAGATCATCATTTAAATAACTTAAATGCTCA 2820
2761 CTTTATTTACACAGTTTATGCTCTGGAATAGATCATCATTTAAATAACTTAAATGCTCA 2820
2821 ACATCATTTGGAGTTATCAGATTTGGAATTAATGTTGAATAACAGTGTCTTATGTAACGCT 2880
2821 ACATCATTTGGAGTTATCAGATTTGGAATTAATGTTGAATAACAGTGTCTTATGTAACGCT 2880
2881 AGCACCAAGTTGTTGATGGGAAACAGGAGCAATATCATTTAGCTTTCAACACACAG 2940
2881 AGCACCAAGTTGTTGATGGGAAACAGGAGCAATATCATTTAGCTTTCAACACACAG 2940
2941 GGAGCACCAAGCTCTCCCAAGATGTTTATTTATGCAAACTCAGTCTTTTCATCAATA 3000
2941 GGAGCACCAAGCTCTCCCAAGATGTTTATTTATGCAAACTCAGTCTTTTCATCAATA 3000
3001 ATTCTTTTCTGAGACACTCTCTTCAAAACCTAATGGGATTTATCAATATTTACTCTGTTAT 3060
3001 ATTCTTTTCTGAGACACTCTCTTCAAAACCTAATGGGATTTATCAATATTTACTCTGTTAT 3060
3061 TACAGAAATACCTTTGAGTACTTTTATGAGAAATTTTACCTCCATGAACTAACCAATGAC 3120
3061 TACAGAAATACCTTTGAGTACTTTTATGAGAAATTTTACCTCCATGAACTAACCAATGAC 3120
3121 TTTGCAATATGACTGTATCCAAATATAGATAAACTGACAAATTTGAGTACTATACA 3180
3121 TTTGCAATATGACTGTATCCCAATATAGATAAACTGACAAATTTGAGTACTATACA 3180
3181 TTTTGGTTAAACAGAAAGTACTTTCAGTTGGAATGGAATGGAAGCAGTGACATCAATTGAA 3240
3181 TTTTGGTTAAACAGAAAGTACTTTCAGTTGGAATGGAATGGAAGCAGTGACATCAATTGAA 3240
3241 GTATACACAGATCAAGACATACCTGAAGGTTTGTGGAACCTGACTTTACGAACTCAT 3300
3241 GTATACACAGATCAAGACATACCTGAAGGTTTGTGGAACCTGACTTTACGAACTCAT 3300
3301 TCGTCAACTGCAATAAATGTAAGCTGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360
3301 TCGTCAACTGCAATAAATGTAAGCTGGGTCCCAACCGGCTCAACCAACCGTCTAGTCTTC 3360
3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCTGCGCATGTGAGACCACTCTTGTGTA 3420
3361 TACTATGTTTCACTGATCTTACAGCAGACTCTCTGCGCATGTGAGACCACTCTTGTGTA 3420
3421 TATGAGAGAGCATATATTTTGAATACTGGAATAACACTGATTTATATATTTAAATAAT 3480
3421 TATGAGAGAGCATATATTTTGAATACTGGAATAACACTGATTTATATATTTAAATAAT 3480
3481 ACTCCATCAACAGAAAGGGAATTTCTGTATACCTATATCTGCCCCAGCTATATCAAGACT 3540

Db	3481		ACTCCA	TC	AA	CAG	AAAA	AGG	AT	TC	TG	TA	AC	CT	TAT	AT	CT	CC	CA	GC	AG	CT	TAT	CA	TA	CA	AG	ACT	3540			
Qy	3541	GA	AG	AG	AT	GT	CC	CA	GA	AA	CT	TC	CA	CA	AT	TA	AT	CA	CA	CT	TT	T	T	T	T	T	T	T	T	3540		
Db	3541	GA	AG	AG	AT	GT	CC	CA	GA	AA	CT	TC	CA	CA	AT	TA	AT	CA	CA	CT	TT	T	T	T	T	T	T	T	T	3500		
Qy	3601	TC	AG	TT	CT	CT	TA	TC	AT	CG	GG	AT	CC	CC	CA	GT	AA	GC	CA	AA	T	GG	T	GC	CA	AT	TA	AT	AG	3660		
Db	3601	TC	AG	TT	CT	CT	TA	TC	AT	CG	GG	AT	CC	CC	CA	GT	AA	GC	CA	AA	T	GG	T	GC	CA	AT	TA	AT	AG	3660		
Qy	3661	TT	AA	CT	TT	TA	CA	AG	CA	CA	AT	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	3720		
Db	3661	TT	AA	CT	TT	TA	CA	AG	CA	CA	AT	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	3720		
Qy	3721	TT	GA	AG	AG	CT	TT	TC	CA	CA	TT	T	AC	AT	T	AC	AT	T	AC	AT	T	AC	AT	T	AC	AT	T	AC	AT	3780		
Db	3721	TT	GA	AG	AG	CT	TT	TC	CA	CA	TT	T	AC	AT	T	AC	AT	T	AC	AT	T	AC	AT	T	AC	AT	T	AC	AT	3780		
Qy	3781	GG	AC	TT	GG	T	CC	TT	CC	AG	TA	T	CT	TT	T	CT	TT	T	AC	AG	AT	GC	AG	T	CA	GC	AG	T	CA	3840		
Db	3781	GG	AC	TT	GG	T	CC	TT	CC	AG	TA	T	CT	TT	T	CT	TT	T	AC	AG	AT	GC	AG	T	CA	GC	AG	T	CA	3840		
Qy	3841	CC	CA	AA	AA	TT	GA	CT	TT	TA	AT	CA	CT	GT	AC	TT	CA	CA	AT	GT	TA	GC	T	GA	AA	AT	GC	AG	CC	CA	3900	
Db	3841	CC	CA	AA	AA	TT	GA	CT	TT	TA	AT	CA	CT	GT	AC	TT	CA	CA	AT	GT	TA	GC	T	GA	AA	AT	GC	AG	CC	CA	3900	
Qy	3901	AG	TC	CT	CT	CC	AG	T	GG	T	AT	GT	T	T	AA	AG	T	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	3960		
Db	3901	AG	TC	CT	CT	CC	AG	T	GG	T	AT	GT	T	AT	AA	AG	T	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	AT	3960		
Qy	3961	GA	CA	CT	AT	AT	T	AT	TA	AG	AA	T	AT	AT	CA	GG	AT	T	T	AA	AA	CT	GA	GC	CA	AA	CT	GT	GG	CA	4020	
Db	3961	GA	CA	CT	AT	AT	T	AT	TA	AG	AA	T	AT	AT	CA	GG	AT	T	T	AA	AA	CT	GA	GC	CA	AA	CT	GT	GG	CA	4020	
Qy	4021	GA	AC	CA	GT	CA	GC	CT	TA	CT	TA	TC	CG	T	GT	AT	CT	GC	GT	TC	CA	AA	AG	TT	GC	CA	AA	CT	GT	GG	CA	4080
Db	4021	GA	AC	CA	GT	CA	GC	CT	TA	CT	TA	TC	CG	T	GT	AT	CT	GC	GT	TC	CA	AA	AG	TT	GC	CA	AA	CT	GT	GG	CA	4080
Qy	4081	CA	AT	TT	AG	TA	AT	GT	AG	T	AA	AA	TT	CA	CA	AA	CC	CA	GA	AT	CA	GT	TC	CA	AG	AT	GC	GT	GC	CA	4140	
Db	4081	CA	AT	TT	AG	TA	AT	GT	AG	T	AA	AA	TT	CA	CA	AA	CC	CA	GA	AT	CA	GT	TC	CA	AG	AT	GC	GT				

4561	DB	CAAAAAATCAATACCTCTATGAAAGCTCACTTAACTGAAGAGACATATATGATGATTAAG	4562
4621	QY	AAATTTAGATGGTATAGATTTCCAAGTGGCTGCCAGCAACCAATGCTGGCTATGGCAATGCT	4680
4621	DB	AAATTTAGATGGTATAGATTTCCAAGTGGCTGCCAGCACCAATGCTGGCTATGGCAATGCT	4680
4681	QY	TCAAACTGGATTTCTACAAAAACTCTGCTCGGCCCTCCAGATGGTCTCTCTGAAAAATGTT	4740
4681	DB	TCAAACTGGATTTCTTACAAAAACTCTGCTCGGCCCTCCAGATGGTCTCTCTGAAAAATGTT	4740
4741	QY	CATGTAGTACCAATCACCCTTTTACATCAGCATTAAGCTGGAGTGAACCTGCTGTCATT	4800
4741	DB	CATGTAGTACCAATCACCCTTTTACATCAGCATTAAGCTGGAGTGAACCTGCTGTCATT	4800
4801	QY	ACTGGACCAACATGTTTATCTGATTTGATGTCAAAATCGGTAGATTAATGATGAATTTAATATA	4860
4801	DB	ACTGGACCAACATGTTTATCTGATTTGATGTCAAAATCGGTAGATTAATGATGAATTTAATATA	4860
4861	QY	TCCTTTCATCAAGTCAAATGAAGAAAAATAAACCATATGAATTTAAGATTTAGAAAAATTTTC	4920
4861	DB	TCCTTTCATCAAGTCAAATGAAGAAAAATAAACCATATGAATTTAAGATTTAGAAAAATTTTC	4920
4921	QY	ACAAGGTATTTCTGTAGTGATCACTGCAATTTACTTGGGACAATTAAGTCTGCATATGTAGAA	4980
4921	DB	ACAAGGTATTTCTGTAGTGATCACTGCAATTTACTTGGGACAATTAAGTCTGCATATGTAGAA	4980
4981	QY	GGGAAGTCAAGTGTGAAATGATTTGTACTTACTTTTAGAATCAGCCCCCAAGGACCCACCT	5040
4981	DB	GGGAAGTCAAGTGTGAAATGATTTGTACTTACTTTTAGAATCAGCCCCCAAGGACCCACCT	5040
5041	QY	AACAACATGATTTTCAGAAAGATACAGATGAAGTTACAAAATTTCAAATTAACGTTTCCCTT	5100
5041	DB	AACAACATGATTTTCAGAAAGATACAGATGAAGTTACAAAATTTCAAATTAACGTTTCCCTT	5100
5101	QY	CCTCCTTCTCAAACCTAATGGAATATCCAAGTATATCAAGCTCTGCTTTACCGAGAAGAT	5160
5101	DB	CCTCCTTCTCAAACCTAATGGAATATCCAAGTATATCAAGCTCTGCTTTACCGAGAAGAT	5160
5161	QY	GATCCTACTGCTGTCCAGATTTCAACAACCTCAGTATTATATACAGAAAAACCAACATTTGCTC	5220
5161	DB	GATCCTACTGCTGTCCAGATTTCAACAACCTCAGTATTATATACAGAAAAACCAACATTTGCTC	5220
5221	QY	ATTGCAATGCTAGAAGGACTTAAAGGTGGACATACATACATATCAGTGTTTTACGCGATC	5280
5221	DB	ATTGCAATGCTAGAAGGACTTAAAGGTGGACATACATACATATCAGTGTTTTACGCGATC	5280
5281	QY	AATAGTCTGCTGCGAGTTCAAAAGTTCGGATCAGATTAACCAATGATATCAAGCTCCA	5340
5281	DB	AATAGTCTGCTGCGAGTTCAAAAGTTCGGATCAGATTAACCAATGATATCAAGCTCCA	5340
5341	QY	GCAGACCAAAAAACCAACCCCTTATTATGATGCCACAGAAAAACTGCTTTGTGACT	5400
5341	DB	GCAGACCAAAAAACCAACCCCTTATTATGATGCCACAGAAAAACTGCTTTGTGACT	5400
5401	QY	TCAACCAAAATTACAATCAGAAATGCCAATATGTTACTACAGTGATGATCATGGACCAATA	5460
5401	DB	TCAACCAAAATTACAATCAGAAATGCCAATATGTTACTACAGTGATGATCATGGACCAATA	5460
5461	QY	AAAAATGTAACAAGTCTTGGCAGAAAAACGAGAGCTCAGCATGATGGAAATGTAAACAAG	5520
5461	DB	AAAAATGTAACAAGTCTTGGCAGAAAAACGAGAGCTCAGCATGATGGAAATGTAAACAAG	5520
5521	QY	TGGTATGATGCATATTTTAAATAAGCAAGGCCATATTTTACAAATGAAGCTTTCTTAAC	5580
5521	DB	TGGTATGATGCATATTTTAAATAAGCAAGGCCATATTTTACAAATGAAGCTTTCTTAAC	5580
5581	QY	CCTCCATGTACAGAGGAAAGACAAAGTTTGTGGCAATGAAGAAATCTACATCATAGGT	5640
5581	DB	CCTCCATGTACAGAGGAAAGACAAAGTTTGTGGCAATGAAGAAATCTACATCATAGGT	5640
5641	QY	GCTGATTAATGCATGCATGATTTCTGGCAATGAAGACAAAAATTTGCAATGGACCACTGAAA	5700
5641	DB	GCTGATTAATGCATGCATGATTTCTGGCAATGAAGACAAAAATTTGCAATGGACCACTGAAA	5700

QY 5701 CCAAAAAGCAATACCTTATTTAAATTTAGAGCTACAAATATTTATGGGACAAATTTACTGAC 5760
DB |||||||
QY 5701 CCAAAAAGCAATACCTTATTTAAATTTAGAGCTACAAATATTTATGGGACAAATTTACTGAC 5760
DB |||||||
QY 5761 TCTGATATTCTGACCTGTGTAAGACTTTAGGGAGGACTTTTCAGAAAAGAACCGGTAGAG 5820
DB |||||||
QY 5761 TCTGATATTCTGACCTGTGTAAGACTTTAGGGAGGACTTTTCAGAAAAGAACCGGTAGAG 5820
DB |||||||
QY 5821 ATCATTTCTTCGGTCACTTTGTGTATCTCTTCAATTAATTTCTCTTGGACAGCTATTTT 5880
DB |||||||
QY 5821 ATCATTTCTTCGGTCACTTTGTGTATCTCTTCAATTAATTTCTCTTGGACAGCTATTTT 5880
DB |||||||
QY 5881 GCATTTGCAAGAAATTCGACAGAGCAGAAAAGGTGGCACATATCTCTCTCAGGATGCA 5940
DB |||||||
QY 5881 GCATTTGCAAGAAATTCGACAGAGCAGAAAAGGTGGCACATATCTCTCTCAGGATGCA 5940
DB |||||||
QY 5941 GAAATTTATGACACTAAATTTGAAGCTTGATCAGCTCATCACAGTGGCAGACCTGGAACCTG 6000
DB |||||||
QY 5941 GAAATTTATGACACTAAATTTGAAGCTTGATCAGCTCATCACAGTGGCAGACCTGGAACCTG 6000
DB |||||||
QY 6001 AAGGACGAGATTAACGGGGCCCAATAAGCAAGAAATCCCTCTCTGCAACATGTTGAAGAG 6060
DB |||||||
QY 6001 AAGGACGAGATTAACGGGGCCCAATAAGCAAGAAATCCCTCTCTGCAACATGTTGAAGAG 6060
DB |||||||
QY 6061 CTTTGCACAAAACAACACTTAAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTT 6120
DB |||||||
QY 6061 CTTTGCACAAAACAACACTTAAAGTTTCAAGAGAAATTTTCGGAATTTACCAAAATTTCTT 6120
DB |||||||
QY 6121 CAGGATCTTTCTTCAACTGATGCTGATCTGCCCTTGAATAGAGCAAAAACCGTTTCCCA 6180
DB |||||||
QY 6121 CAGGATCTTTCTTCAACTGATGCTGATCTGCCCTTGAATAGAGCAAAAACCGTTTCCCA 6180
DB |||||||
QY 6181 AACATTAACCATATATTAATTAATTAACAGATTAAGCTGATGACGCTAGTGTTC 6240
DB |||||||
QY 6181 AACATTAACCATATATTAATTAATTAACAGATTAAGCTGATGACGCTAGTGTTC 6240
DB |||||||
QY 6241 GGTTCGGATTATTAATTAATGCGAGCTATATTTCTCGTTTATTTATGTCCTCAATGAATTTAT 6300
DB |||||||
QY 6241 GGTTCGGATTATTAATTAATGCGAGCTATATTTCTCGTTTATTTATGTCCTCAATGAATTTAT 6300
DB |||||||
QY 6301 GCTACTCAAGGTCCTACTACCAGGAAACAGTTGGAGATTTTGGAGATGTTGGGAAACC 6360
DB |||||||
QY 6301 GCTACTCAAGGTCCTACTACCAGGAAACAGTTGGAGATTTTGGAGATGTTGGGAAACC 6360
DB |||||||
QY 6361 AGGGCAAAAACATTAGTAATGCTAACACAGTGTGTTTGAAGGACGGATCAGATGCCAT 6420
DB |||||||
QY 6361 AGGGCAAAAACATTAGTAATGCTAACACAGTGTGTTTGAAGGACGGATCAGATGCCAT 6420
DB |||||||
QY 6421 CAGTATTTGGCCAGAGACCAACAGCCAGTTTCTCTTTGGAGATATAGTGAATTAACAAG 6480
DB |||||||
QY 6421 CAGTATTTGGCCAGAGACCAACAGCCAGTTTCTCTTTGGAGATATAGTGAATTAACAAG 6480
DB |||||||
QY 6481 CTAATGGAGGATGTTCAAAATAGATTGCACTATCAGGGATCTGAAAATTTGAAAGGCGATGGG 6540
DB |||||||
QY 6481 CTAATGGAGGATGTTCAAAATAGATTGCACTATCAGGGATCTGAAAATTTGAAAGGCGATGGG 6540
DB |||||||
QY 6541 GATTGATGATGTTTCGACAGTGTAACTTTA CTGCTGCGCCAGAGATGGGTTTCTGAG 6600
DB |||||||
QY 6541 GATTGATGATGTTTCGACAGTGTAACTTTA CTGCTGCGCCAGAGATGGGTTTCTGAG 6600
DB |||||||
QY 6601 AACAGGCCCCCTTAATTTCACTTTGTGAAGTTGGTTTCGAGCAGCAAGGGGCATGACACC 6660
DB |||||||
QY 6601 AACAGGCCCCCTTAATTTCACTTTGTGAAGTTGGTTTCGAGCAGCAAGGGGCATGACACC 6660
DB |||||||
QY 6661 ACACCTATGATTTTCACTGCACTGCTGGAGTTGGAAGAACTGGAGTTTTTATTTGCTCTG 6720
DB |||||||
QY 6661 ACACCTATGATTTTCACTGCACTGCTGGAGTTGGAAGAACTGGAGTTTTTATTTGCTCTG 6720
DB |||||||
QY 6721 GACCATTAAACACATATTAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCT 6780
DB |||||||
QY 6721 GACCATTAAACACATATTAATGACCATGATTTTGTGGATATATATGAGCTAGTAGCT 6780
DB |||||||

QY 6781 GAACTGAGAAGTGAAGAATGTGCATGTGCGAAGATCTGGCAGCAGTATATCTTTTACAC 6840
DB |||||||
QY 6781 GAACTGAGAAGTGAAGAATGTGCATGTGCGAAGATCTGGCAGCAGTATATCTTTTACAC 6840
DB |||||||
QY 6841 CAGTGCATTTCTGGATCTCTTATCAAAATAGGGAAGTAAATCAGCCCATCTGTTTGTAAAC 6900
DB |||||||
QY 6841 CAGTGCATTTCTGGATCTCTTATCAAAATAGGGAAGTAAATCAGCCCATCTGTTTGTAAAC 6900
DB |||||||
QY 6901 TATTTCAGCACTTTCAGAAAGTGGACTCTTTTGGACGCCATGGAAAGGTGATGTTTCAGCTTGAA 6960
DB |||||||
QY 6901 TATTTCAGCACTTTCAGAAAGTGGACTCTTTTGGACGCCATGGAAAGGTGATGTTTCAGCTTGAA 6960
DB |||||||
QY 6961 TGGGAAGAAAACCACTATGTAAATATTCAGACCCAAAGGATACAAATTCGGAAGAGATTTTAA 7020
DB |||||||
QY 6961 TGGGAAGAAAACCACTATGTAAATATTCAGACCCAAAGGATACAAATTCGGAAGAGATTTTAA 7020
DB |||||||
QY 7021 ATCCAGAGGGGCCAAAAGTTACCCCTCATTTCCGAAATGAAATGTGCACCTTAAAGAA 7080
DB |||||||
QY 7021 ATCCAGAGGGGCCAAAAGTTACCCCTCATTTCCGAAATGAAATGTGCACCTTAAAGAA 7080
DB |||||||
QY 7081 ATATCTATGCTTCTCTCACTGTCCTTT 7108
DB |||||||
QY 7081 ATATCTATGCTTCTCTCACTGTCCTTT 7108
DB |||||||

RESULT 3
US-10-466-759-10
; Sequence 10, Application US/10466759
; Publication No. US20040081983A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: LEE, Ernestine A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: ISON, Craig H.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: TRAN, Bao
; APPLICANT: DING, Li
; APPLICANT: LU, Dying Aina M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: WARREN, Bridget A.
; TITLE OF INVENTION: KINASES AND PHOSPHATASES
; FILE REFERENCE: PI-0344 USN
; CURRENT APPLICATION NUMBER: US/10/466,759
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: PCT/US02/01369
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: US 60/263,083
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/271,205
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/271,117
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/276,859
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/278,504
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/278,522
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/280,510
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/280,266
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PERL Program
; SEQ ID NO 10
; LENGTH: 7225
; TYPE: DNA

Qy	2128	ATTGATGTTACCGCAGATGAATAATGAGTTGAGTGGTTCACCA	CCGCGAAAGCCCAATGGG	2181
Db	2023	ATTGATGTTTACCGCAGATGAATAATGAGTTGAGTGGTTCACCA	CCGCGAAAGCCCAATGGG	2082
Qy	2188	ATCAATTATTGCTTATGAAGTGCTATATAAAATATAGATATCTT	TATATATATGAAGAACACA	2247
Db	2083	ATCAATTATTGCTTATGAAGTGCTATATAAAATATAGATATCTT	TATATATATGAAGAACACA	2142
Qy	2248	TCAACAACAGACATAAATATTAAAGGAACTTAAAGACCTTCAC	CCCTCTATAAATTTCTGTGA	2307
Db	2143	TCAACAACAGACATAAATATTAAAGGAACTTAAAGACCTTCAC	CCCTCTATAAATTTCTGTGA	2202
Qy	2308	AGGTCCTTACACAGATTTGGTTCATGCGCAATCAGGTATCTCT	TTTTCCTCTGTGAAGGACT	2367
Db	2203	AGGTCCTTACACAGATTTGGTTCATGCGCAATCAGGTATCTCT	TTTTCCTCTGTGAAGGACT	2262
Qy	2368	TCGAGAGCTGCGCTGATAGTGACACAGAAAATCACTTACAAA	ATATTTCTTCGGA	2427
Db	2263	TCGAGAGCTGCGCTGATAGTGACACAGAAAATCACTTACAAA	ATATTTCTTCGGA	2322
Qy	2428	GAGATTGAGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGA	ATCATPAAAAATATTTCTTCGGA	2487
Db	2323	GAGATTGAGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGA	ATCATPAAAAATATTTCTTCGGA	2382
Qy	2488	ATTTATCTCAAGAGAAGTAATGGAATGAGGAAAGAACTATA	ATAATACAACTCTTTAAAC	2547
Db	2383	ATTTATCTCAAGAGAAGTAATGGAATGAGGAAAGAACTATA	ATAATACAACTCTTTAAAC	2442
Qy	2548	CAAAACATTAAAGTACTGAGAAATATACCAATATATCATTG	AGGTGCTGCTAGTAC	2607
Db	2443	CAAAACATTAAAGGACTGAGAAATATACCCATATATCATTG	AGGTGCTGCTAGTAC	2502
Qy	2608	CTGAAAGGTGAAGGAGTTTCGGAGTGCTCCCATAGTATATC	GCACGGAGGAAGATGCTCCT	2667
Db	2503	CTCAAGAGTGAAGGAGTTTCGGAGTGCTCCCATAGTATATC	GCACGGAGGAAGATGCTCCT	2562
Qy	2668	GATTCTCCCCCTCAAGACTTCTCTGTAAAAACAGTTGTCT	GGTGTCACGGTGAATGTGCA	2727
Db	2563	GATTCTCCCCCTCAAGACTTCTCTGTAAAAACAGTTGTCT	GGTGTCACGGTGAATGTGCA	2622
Qy	2728	TGGCAACACCCCTGGAGCCCAATGGAATTTATCCTTTATT	TACACAGTTTATGCT---GG	2784
Db	2623	TGGCAACACCCCTGGAGCCCAATGGAATTTATCCTTTATT	TACACAGTTTATGCTGGAGG	2682
Qy	2785	AATAGATCATCATTTAAAAACTATTAAATGTCACTGAAA	CATCATTTGGAGTTTATCAGATTG	2844
Db	2683	AATAGATCATCATTTAAAAACTATTAAATGTCACTGAAA	CATCATTTGGAGTTTATCAGATTG	2742
Qy	2845	GATTATAATGTTGAATACAGTGCTTATGTAAACAGCTAG	CACCAAGATTGGTGATGGGAAA	2904
Db	2743	GATTATAATGTTGAATACAGTGCTTATGTAAACAGCTAG	CACCAAGATTGGTGATGGGAAA	2802
Qy	2905	ACAGGAAGCAATATCATTTAGCTTTCAAAACACGAGG	GAGGACCAAGCGATCTCCCAAA	2964
Db	2803	ACAGGAAGCAATATCATTTAGCTTTCAAAACACGAGG	GAGGACCAAGCGATCTCCCAAA	2859
Qy	2965	GATGTTTATTATGCAAACTCAGTTCTTCATCAATAAT	CTTTTCTGCGACACCTCCTTCA	3024
Db	2860	GATGTTTATTATGCAAACTCAGTTCTTCATCAATAAT	CTTTTCTGCGACACCTCCTTCA	2919
Qy	3025	AAACCTAATGGGATATACAATATTACTCTGTTTATTAC	GAGAAATACTTCAGGTACTTTTT	3084
Db	2920	AAACCTAATGGGATATACAATATTACTCTGTTTATTAC	GAGAAATACTTCAGGTACTTTTT	2979
Qy	3085	ATGCAGAAATTTTACACTCCATGAACTAAACCAATGA	CTTTTGGACAAATATGACTGTATCCACA	3144
Db	2980	ATGCAGAAATTTTACACTCCATGAACTAAACCAATGA	CTTTTGGACAAATATGACTGTATCCACA	3039
Qy	3145	ATTATAGATAAACTGCAAAATTTTACAGCTACTATACA	TTTTTGGTTTAAACAGCAAGTACTTCA	3204
Db	3040	ATTATAGATAAACTGCAAAATTTTACAGCTACTATACA	TTTTTGGTTTAAACAGCAAGTACTTCA	3099
Qy	3205	GTTTGGAAATGGGAATAAAGACAGTGACATCATTTGA	AGTATATACAGATCAAGACATACCT	3264

[illegible]

Db 601 GTAGTGAAGAGTCTCAATCAGAGTAGAGGACATTTTCACCTGGGAAATTTGCAGAAATGC 660
Qy 763 AAT--GAGAATAGTGAATCTTTTATGGAGTACAGCCGCTTCTCAACCCCTTGGT 819
Db 661 AATGTAGAGATAGTGAATCTTTTATGGAGTACAGCCGCTTCTCAACCCCTTGGT 720
Qy 820 AGAGTTACACCTCCATCGGCTGACCAACAATTCATCAAGCAAGTGTGACACAGAAATGATC 879
Db 721 AGAGTTACACCTCCATCGGCTGACCAACAATTCATCAAGCAAGTGTGACACAGAAATGATC 780
Qy 880 AGCTCTGTGTGAAGAGAGCTATCAGTTTGTGTGACACACTTGTGAGACCTTATACAACA 939
Db 781 AGCTCTGTGTGAAGAGAGCTATCAGTTTGTGTGACACACTTGTGAGACCTTATACAACA 840
Qy 940 TATCTTTTGAAGTTTTCAGCTGCTACACTGAGCAGAGTTTATTTGATAGTACGATTTGC 999
Db 841 TATCTTTTGAAGTTTTCAGCTGCTACACTGAGCAGAGTTTATTTGATAGTACGATTTGC 900
Qy 1000 AGAACACAGAAATCAGTGTGAGGACCAACAACAACCTGCGTAAACAGGCAACATCAACA 1059
Db 901 AGAACACAGAAATCAGTGTGAGGACCAACAACAACCTGCGTAAACAGGCAACATCAACA 960
Qy 1060 GGAAGTCTCTTTCAATTTTATGGGACCCACCAACTATAGTAAACAGGGAATTTAGTTAT 1119
Db 961 GGAAGTCTCTTTCAATTTTATGGGACCCACCAACTATAGTAAACAGGGAATTTAGTTAT 1020
Qy 1120 AGAGTTGAATTTATGGAACCATCAGCTGCAATTTTGGATAAAGCAACCAAGACCTCAAG 1179
Db 1021 AGAGTTGAATTTATGGAACCATCAGCTGCAATTTTGGATAAAGCAACCAAGACCTCAAG 1080
Qy 1180 TTTCATTTCACTAACTAACAACATTTAATGATGATGATGATGATGATGATGATGATGATGAT 1239
Db 1081 TTTCATTTCACTAACTAACAACATTTAATGATGATGATGATGATGATGATGATGATGATGAT 1140
Qy 1240 AGTCAGGAGACTGGGCCCCAAGTCAATATTTTCAGTATTTCACTCCACCAAGTGTCCAGGG 1299
Db 1141 AGTCAGGAGACTGGGCCCCAAGTCAATATTTTCAGTATTTCACTCCACCAAGTGTCCAGGG 1200
Qy 1300 GCAGTGTGTTGATTTACAACTTGACAGAGTAGAATCCACCAAGTAAAGTAAATTTCTGGAAG 1359
Db 1201 GCAGTGTGTTGATTTACAACTTGACAGAGTAGAATCCACCAAGTAAAGTAAATTTCTGGAAG 1260
Qy 1360 AAACCAACGACCAACCAATGGAATTTAATCAATGATGATGATGATGATGATGATGATGATGAT 1419
Db 1261 AAACCAACGACCAACCAATGGAATTTAATCAATGATGATGATGATGATGATGATGATGATGAT 1320
Qy 1420 ACAGGAATATTTTGGAAATTTCTTCTCACTGGAATTAATGAGTATTAATGACCC 1479
Db 1321 ACAGGAATATTTTGGAAATTTCTTCTCACTGGAATTAATGAGTATTAATGACCC 1377
Qy 1480 ATGGCTCCAGAAATTTGGAACATAGTAGAGCCAAATGTTAGGATTTATATGAGGTTTCAGCA 1539
Db 1378 ATGGCTCCAGAAATTTGGAACATAGTAGAGCCAAATGTTAGGATTTATATGAGGTTTCAGCA 1437
Qy 1540 GAGATGTCTGTGACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1599
Db 1438 GAGATGTCTGTGACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497
Qy 1600 AACTTCTCTGGAAGAAATGAGCTGGAAGCAAGCTTCACTGGAATTAATGAGTATTAATGAGGAAAT 1659
Db 1498 AACTTCTCTGGAAGAAATGAGCTGGAAGCAAGCTTCACTGGAATTAATGAGTATTAATGAGGAAAT 1557
Qy 1660 CAGTATATTACTGACATTCAGCTGGAAGCAAGCTTCACTGGAATTAATGAGTATTAATGAGGAAAT 1719
Db 1558 CAGTATATTACTGACATTCAGCTGGAAGCAAGCTTCACTGGAATTAATGAGTATTAATGAGGAAAT 1617
Qy 1720 TTTCATCTGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1779
Db 1618 TTTCATCTGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677
Qy 1780 GTTCTCAGTGTGAGGACAGTCAAGAGTCCAGCTCAATTTAAATTTAAATTTAAATTTAAATTTAA 1839

Db 1678 GTTCTCAGTGTGAGGACAGCTCAGCAAGTCCAGAGCTCAATTTAAATTTAAATTTAAATTTAA 1737
Qy 1840 AATATTAGTCTTCTCATCTATTTTGTATATGGGATCTCTCCAGAAATATCCCAATGGAAAA 1899
Db 1738 AATATTAGTCTTCTCATCTATTTTGTATATGGGATCTCTCCAGAAATATCCCAATGGAAAA 1797
Qy 1900 AATACCTCACTATACATTTATGCAATGGAATTTGGATACAAACAGAGCAATTCAGATAACT 1959
Db 1798 AATACCTCACTATACATTTATGCAATGGAATTTGGATACAAACAGAGCAATTCAGATAACT 1857
Qy 1960 ACCATAGATAACAGCTTTCTCATTAAC-----AGGTTTAAAGAAATATACACAAATACAAA 2013
Db 1858 ACCATAGATAACAGCTTTCTCATACAGGTATAGGTTTAAAGAAATATACACAAATACAAA 1917
Qy 2014 ATGAGAGTGGCAGCTTCAACCCAGATGGAGAAAGTTCTTTCTGTGAAGAAATGACATC 2073
Db 1918 ATGAGAGTGGCAGCTTCAACCCAGCTTGGAGAAAGTTCTTTCTGTGAAGAAATGACATC 1977
Qy 2074 TTTGTGAGAACTTCAAGAGATGAACCGGAATCATCACCTCAAGATGTGCAAGTAAATGAT 2133
Db 1978 TTTGTGAGAACTTCAAGAGATGAACCGGAATCATCACCTCAAGATGTGCAAGTAAATGAT 2037
Qy 2134 GTTACCGCAGATGAATAAGGTTGAAGTGTCCACCCGAAAGCCCAATGSGGATCAT 2193
Db 2038 GTTACCGCAGATGAATAAGGTTGAAGTGTCCACCCGAAAGCCCAATGSGGATCAT 2097
Qy 2194 ATTTGCTTATGAAGTCTATATATAAATATAGATCTTTATATATGAAGAAACATCAACA 2253
Db 2098 ATTTGCTTATGAAGTCTATATATAAATATAGATCTTTATATATGAAGAAACATCAACA 2157
Qy 2254 ACAGACATTAATTAAGGAACCTTAAGACCTCAACCTCTATAACAATTTCTGTAAGGTCT 2313
Db 2158 ACAGACATTAATTAAGGAACCTTAAGACCTCAACCTCTATAACAATTTCTGTAAGGTCT 2217
Qy 2314 TACACAGATTTGTTGTCATGGCAATCAGGTATCTTCTTACTCTCTGTAAGGACTTCGGAG 2373
Db 2218 TACACAGATTTGTTGTCATGGCAATCAGGTATCTTCTTACTCTCTGTAAGGACTTCGGAG 2277
Qy 2374 ACTGTGCTCTGATAGTGCACCAAGAAATATCACTTACAAATAATTTCTTCTGGAGAGATT 2433
Db 2278 ACTGTGCTCTGATAGTGCACCAAGAAATATCACTTACAAATAATTTCTTCTGGAGAGATT 2337
Qy 2434 GAGCTATCAATTCCTTCCCAAGTATGCCAATGGAATCATAAAAATATATCAATTTAT 2493
Db 2338 GAGCTATCAATTCCTTCCCAAGTATGCCAATGGAATCATAAAAATATATCAATTTAT 2397
Qy 2494 CTCAGAGAGTAAATGGAATGAGAAAGAACTATAAATACAACTTTTAAACCCAAAC 2553
Db 2398 CTCAGAGAGTAAATGGAATGAGAAAGAACTATAAATACAACTTTTAAACCCAAAC 2457
Qy 2554 ATTAAGATTACTGAAGAAATATACCAATATATCACTTGAAGTCTCTGCTAGTACACTGAAA 2613
Db 2458 ATT-----CTGAAGAAATATACCAATATATCACTTGAAGTCTCTGCTAGTACACTGAAA 2511
Qy 2614 GGTGAAGAGTTTGGAGTCTCCCAATAGTATATCTGAAGGAGAAAGATGCTCTGATTTCT 2673
Db 2512 GGTGAAGAGTTTGGAGTCTCCCAATAGTATATCTGAAGGAGAAAGATGCTCTGATTTCT 2571
Qy 2674 CCCCTCTCAAGACTTCTCTGTAATAACAGTTGTCTGCTGTCACCGTGAAGTTGTTCATGGCAA 2733
Db 2572 CCCCTCTCAAGACTTCTCTGTAATAACAGTTGTCTGCTGTCACCGTGAAGTTGTTCATGGCAA 2631
Qy 2734 CCACCCCTGGAGCCAAATGGAATTTATCTTTATATACAGTTTATGTTCT--GGAATAGA 2790
Db 2632 CCACCCCTGGAGCCAAATGGAATTTATCTTTATATACAGTTTATGTTCTGGAGGAAATAGA 2691
Qy 2791 TCATCATTTAAAACTATTAATGTCTGAACATCATTTGGAGTTATCAAGTTTGGATTAT 2850
Db 2692 TCATCATTTAAAACTATTAATGTCTGAAACATCATTTGGAGTTATCAAGTTTGGATTAT 2751
Qy 2851 AATGTTGAATACAGTCTTATGTAACAGCTAGCACAGATTTGGTGTGGGAAACAGGA 2910
Db 2752 AATGTTGAATACAGTCTTATGTAACAGCTAGCACAGATTTGGTGTGGGAAACAGGA 2811

QY	2911	AGCAATATCATTAGCTTTTCAAACACCCAGAGGGAGCACCAAGCGATCCTCCCAAGATGTT	2970
DB	2812	AGCAATATCATTAGCTTTCAAACACCCAGAGGG---ACCAAGCGATCTCCCAAGATGTT	2868
QY	2971	TATTATGCAAAACCTCAGTTCTTTCATCAATAAATTTCTTTTCTGGGACACCTCTCTCAAAACCT	3030
DB	2869	TATTATGCAAAACCTCAGTTCTTTCATCAATAAATTTCTTTTCTGGGACACCTCTCTCAAAACCT	2928
QY	3031	AATGGGATATACAAATAATTAATCTCTGTTTATTACAGAAATATCTTCAGGTACTTTTATGCGAG	3090
DB	2929	AATGGGATATACAAATAATTAATCTCTGTTTATTACAGAAATATCTTCAGGTACTTTTATGCGAG	2988
QY	3091	AAATTTTACATCCCACTGAATTAACCAATGACTTTTGCACAAATATGACTGTATCCACAATATATA	3150
DB	2989	AAATTTTACATCCCACTGAATTAACCAATGACTTTTGCACAAATATGACTGTATCCACAATATATA	3048
QY	3151	GATAAACTGACAAATATTACAGCTACTATACTATTTTGGTTAAACAGCAAGTACTTCAGTTGGGA	3210
DB	3049	GATAAACTGACAAATATTACAGCTACTATACTATTTTGGTTAAACAGCAAGTACTTCAGTTGGGA	3108
QY	3211	AATGGGAATAAAGCAGTGACATCAATTTGAGTATACACAGATCAAGACATACCTTGAGGG	3270
DB	3109	AATGGGAATAAAGCAGTGACATCAATTTGAGTATACACAGATCAAGACATACCTTGAGGG	3168
QY	3271	TTTGTTTGGAAACCTGACTTACGAATCCAAATTCGTTCAACTGCGAATAAATGTAAGCTGGGTC	3330
DB	3169	TTTGTTTGGAAACCTGACTTACGAATCCAAATTCGTTCAACTGCGAATAAATGTAAGCTGGGTC	3228
QY	3331	CCAACGGCTCAACCAAAACGGTCTAGTCTTCTAATGTTTCACTGATCTTTACACGACACT	3390
DB	3229	CCAACGGCTCAACCAAAACGGTCTAGTCTTCTAATGTTTCACTGATCTTTACACGACACT	3288
QY	3391	CCTGCCATGTGAGACCAACCTCTTGTTCATATACATGACAGAGACATATATTTTGATAATCTG	3450
DB	3289	CCTGCCATGTGAGACCAACCTCTTGTTCATATACATGACAGAGACATATATTTTGATAATCTG	3348
QY	3451	GAAAAATACACTGATATATATAATTAATAAATAATTAATCCATCAACAGAAAAAGGAGTCTCTGAT	3510
DB	3349	GAAAAATACACTGATATATATAATTAATAAATAATTAATCCATCAACAGAAAAAGGAGTCTCTGAT	3408
QY	3511	ACCTATACCTGCCAGCTATACATCAAGACTGAAGAAGATGTCACAGAAACTTCCACCAATA	3570
DB	3409	ACCTATACCTGCCAGCTATACATCAAGACTGAAGAAGATGTCACAGAAACTTCCACCAATA	3468
QY	3571	ATCAACACTTTTAAAAACCTTTTCTTACCTCAGTTCTTATCATCGGGATCCCCAGTA	3630
DB	3469	ATCAACACTTTTAAAAACCTTTTCTTACCTCAGTTCTTATCATCGGGATCCCCAGTA	3528
QY	3631	AAGCCAAATGCTGCAATAAATGAATTAATGATTTAACTTTTACAAGGACCAAAATGAAATTTAT	3690
DB	3529	AAGCCAAATGCTGCAATAAATGAATTAATGATTTAACTTTTACAAGGACCAAAATGAAATTTAT	3588
QY	3691	TCCTTTCACTTCTCTGATAATTAACAATAATTTGGAAGAGCTTTTCAACCAATTTACATTTATAT	3750
DB	3589	TCCTTTCACTTCTCTGATAATTAACAATAATTTGGAAGAGCTTTTCAACCAATTTACATTTATAT	3648
QY	3751	AGCTTTTGTGTCGCGCAAGAACTAGAAAAAGGACTTGGTCTCTTCCAGTATCTTTTCTTTT	3810
DB	3649	AGCTTTTGTGTCGCGCAAGAACTAGAAAAAGGACTTGGTCTCTTCCAGTATCTTTTCTTTT	3708
QY	3811	TACACAGATGATCAGTGCCGTGAGCACTTCCACAAATTTGACTTTTAAATCACTGTACT	3870
DB	3709	TACACAGATGATCAGTGCCGTGAGCACTTCCACAAATTTGACTTTTAAATCACTGTACT	3768
QY	3871	TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA	3930
DB	3769	TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTCCAGGTGGTATTTGTTAAAGTA	3828
QY	3931	TATAGTTTTAAAAATTCAATGAACATGAACCTGACACTATATATATAAGATATATACAGGA	3990
DB	3829	TATAGTTTTAAAAATTCAATGAACATGAACCTGACACTATATATATAAGATATATACAGGA	3888

[illegible]

Db 1498 AACTTCTCTGCAAGGAATAGAGCTGAAGACCGAGCTTCCAGCTGTGTAACTCAAGGAAT 1557
Qy 1660 CAGTATATTAATGACATTTGACGTGAACAGCTGTCTTATGTATTCAGGAGACTTGTACCT 1719
Db 1558 CAGTATATTAATGACATTTGACGTGAACAGCTGTCTTATGTATTCAGGAGACTTGTACCT 1617
Qy 1720 TTCACTGAGCAGATGATTAGTGTATCTGCTTTTACCATCATGCGAGAGGACACCAACA 1779
Db 1618 TTCACTGAGCAGATGATTAGTGTATCTGCTTTTACCATCATGCGAGAGGACACCAACA 1677
Qy 1780 GTTCTCAGTGTATAGGACAGCTCAGCAAGTGCCTCAAGCTCCATTAATAATTAATAACTATAAA 1839
Db 1678 GTTCTCAGTGTATAGGACAGCTCAGCAAGTGCCTCAAGCTCCATTAATAATTAATAACTATAAA 1737
Qy 1840 AATATTAGTTCCTTCATCTAATTTTGTATATTTGGATCTCTCAGAAATATCCCAATGGAAAA 1899
Db 1738 AATATTAGTTCCTTCATCTAATTTTGTATATTTGGATCTCTCAGAAATATCCCAATGGAAAA 1797
Qy 1900 ATAACTCATTATACGATTTATGCAATGGAAATTTGGATATCAAAACAGAGATTCAGATACT 1959
Db 1798 ATAACTCATTATACGATTTATGCAATGGAAATTTGGATATCAAAACAGAGCATTCAGATACT 1857
Qy 1960 ACCATAGATAACAGCTTTCTCATAAC-----AGGGTTAAAGAAATACACAAATACAAA 2013
Db 1858 ACCATAGATAACAGCTTTCTCATAACAGGTATAGGGTTAAAGAAATACACAAATACAAA 1917
Qy 2014 ATGAGAGTGCAGCTCAACCCAGATGGAGAAAGTTCTTTGTCTGAAGAAATGACATC 2073
Db 1918 ATGAGAGTGCAGCTCAACCCAGTTCGGAAGTTCTTTGTCTGAAGAAATGACATC 1977
Qy 2074 TTTGTGAGAACTTCAGAGATGAACCGGAATCATCACTCAAGATGTGGAAGTAAATGAT 2133
Db 1978 TTTGTGAGAACTTCAGAGATGAACCGGAATCATCACTCAAGATGTGGAAGTAAATGAT 2037
Qy 2134 GTTACCGCAGATCAAAATAGTTGAAGTGTGACCAACCGGAAGCCCAATGGATCAT 2193
Db 2038 GTTACCGCAGATCAAAATAGTTGAAGTGTGACCAACCGGAAGCCCAATGGATCAT 2097
Qy 2194 ATTGCTTATGAAGTGTATATAAAATATAGATATCTTTATATATGAAGAAACATCAACA 2253
Db 2098 ATTGCTTATGAAGTGTATATAAAATATAGATATCTTTATATATGAAGAAACATCAACA 2157
Qy 2254 ACAGACATATAATTAAGGAACCTTAAGACCTCAACCCCTCTATTAACATTTCTGTAAGGTCT 2313
Db 2158 ACAGACATATAATTAAGGAACCTTAAGACCTCAACCCCTCTATTAACATTTCTGTAAGGTCT 2217
Qy 2314 TACACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG 2373
Db 2218 TACACAGATTTGGTCAATGCAATCAGGTATCTTTTACTCTCTGTAAGGACTTCGGAG 2277
Qy 2374 ACTGTGCTGATAGTGACACAGAAATATCACTTACAAAAATATTTCTCTGAGAGATT 2433
Db 2278 ACTGTGCTGATAGTGACACAGAAATATCACTTACAAAAATATTTCTCTGAGAGATT 2337
Qy 2434 GAGCTATCATTTCTCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTAT 2493
Db 2338 GAGCTATCATTTCTCCCAAGTAGTCCCAATGGAATCATAAAAAATATACAAATTTAT 2397
Qy 2494 CTCAGAGAGATTAATGGAATGAGGAAGAACTATATAATACAACTCTTTTAAACCCCAAAAC 2553
Db 2398 CTCAGAGAGATTAATGGAATGAGGAAGAACTATATAATACAACTCTTTTAAACCCCAAAAC 2457
Qy 2554 ATTTAAAGTACTGAAGAAATATACCAATATATCATTTAGGTGTCTGTAGTACACTGAAA 2613
Db 2458 ATT-----CTGAAGAAATATACCAATATATCATTTAGGTGTCTGTAGTACACTCAAA 2511
Qy 2614 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACCGGAGGAAGTCTCTGATTCT 2673
Db 2512 GGTGAAGGAGTTCGGAGTGTCTCCATAAGTATATCTGACCGGAGGAAGTCTCTGATTCT 2571
Qy 2674 CCCCCCTCAAGACTTCTCTGTAAACAGATTGTCTGGTGTCAACCGTGAAGTTGTGATGGCAA 2733
Db 2572 CCCCCCTCAAGACTTCTCTGTAAACAGATTGTCTGGTGTCAACCGTGAAGTTGTGATGGCAA 2631

Qy 2734 CCACCCCTGGAGCCAAATGGAATTAATCTTTTATTACAGTTTATGTCT---GGAATAGA 2790
Db 2632 CCACCCCTGGAGCCAAATGGAATTAATCTTTTATTACAGTTTATGTCTCTGGAGGAATAGA 2691
Qy 2791 TCATCATTAATAAACTATTAAATGTCTCACTGAAACATCATCTGGAGTTATCAGATTTGGATTAT 2850
Db 2692 TCATCATTAATAAACTATTAAATGTCTCACTGAAACATCATCTGGAGTTATCAGATTTGGATTAT 2751
Qy 2851 AATGTTGAATACAGTCTCTTATGTAAACAGCTAGCACCAGATTTGGTGTATGGAAAAACAGGA 2910
Db 2752 AATGTTGAATACAGTCTCTTATGTAAACAGCTAGCACCAGATTTGGTGTATGGAAAAACAGGA 2811
Qy 2911 AGCAATATCATTTAGCTTTCAAAACAGAGGAGGACCAAGCGATCCTCCCAAGATGTT 2970
Db 2812 AGCAATATCATTTAGCTTTCAAAACAGAGGAGG---ACCAAGCGATCCTCCCAAGATGTT 2868
Qy 2971 TATTATGCAAACTCAGTTCTTTCATCAATAATTTCTTTCTGACACCTCTCTTCAAAACCT 3030
Db 2869 TATTATGCAAACTCAGTTCTTTCATCAATAATTTCTTTCTGACACCTCTCTTCAAAACCT 2928
Qy 3031 AATGGATTAATAAAATTTACTCTGTATTATACAGAAATACCTTCAGTACTTTTATGTCAG 3090
Db 2929 AATGGATTAATAAAATTTACTCTGTATTATACAGAAATACCTTCAGTACTTTTATGTCAG 2988
Qy 3091 AATTTTACACTCCATGAACCTAACCAATGACTTTTGAACAATATGATGTATCCCAATATATA 3150
Db 2989 AATTTTACACTCCATGAAGTAACTAATGACTTTTGAACAATATGATGTATCCCAATATATA 3048
Qy 3151 GATTAACCTGACATATTTAGCTACTATATATTTTGGTTAAACAGCAAGTACTTCAGTTGGA 3210
Db 3049 GATTAACCTGACATATTTAGCTACTATATATTTTGGTTAAACAGCAAGTACTTCAGTTGGA 3108
Qy 3211 AATGGGAATTAAGCAGTGCACATCATTTGAAGTATACACAGATCAACAGACATCTGGAAGG 3270
Db 3109 AATGGGAATTAAGCAGTGCACATCATTTGAAGTATACACAGATCAACAGACATCTGGAAGG 3168
Qy 3271 TTTGTTGGAACCTGACTTTACGAATCCATTTCTGTCACCTGCAATTAATTAAGTCTGGTCT 3330
Db 3169 TTTGTTGGAACCTGACTTTACGAATCCATTTCTGTCACCTGCAATTAATTAAGTCTGGTCT 3228
Qy 3331 CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACT 3390
Db 3229 CCACCGGCTCAACCAAAACGGTCTAGTCTTCTACTATGTTTCTACTGATCTTACAGCAGACT 3288
Qy 3391 CTCTGCCATGTCAGACCACTCTCTGTTACATATGAGAGAGCAATATTTTGTGATAATCTG 3450
Db 3289 CTCTGCCATGTCAGACCACTCTCTGTTACATATGAGAGAGCAATATTTTGTGATAATCTG 3348
Qy 3451 GAAAAATACACTGATTATATATTAATAAAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3510
Db 3349 GAAAAATACACTGATTATATATTAATAAAATTTACTCCATCAACAGAAAGGATTTCTCTGAT 3408
Qy 3511 ACCTATATGCCCCAGCTATATACATCAAGACTGAAGAGATGTCCCGAGAACTTCAACATA 3570
Db 3409 ACCTATATGCCCCAGCTATATACATCAAGACTGAAGAGATGTCCCGAGAACTTCAACATA 3468
Qy 3571 ATCAACACTTTTAAAAACCTTTCTCTACTCTCAGTCTCTTATCATTTGGATCCCCAGTA 3630
Db 3469 ATCAACACTTTTAAAAACCTTTCTCTACTCTCAGTCTCTTATCATTTGGATCCCCAGTA 3528
Qy 3631 AAGCCAAATGTTGCAATTAATAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAAATTAT 3690
Db 3529 AAGCCAAATGTTGCAATTAATAGTTATGATTTAACTTTTCAAGGACCAAAATGAAAAATTAT 3588
Qy 3691 TCTTTTCACTTCTGATAATTAATAATATTTGGAAGAGCTTTCACCAATTAATATAT 3750
Db 3589 TCTTTTCACTTCTGATAATTAATAATATTTGGAAGAGCTTTCACCAATTAATATAT 3648
Qy 3751 AGCTTTTGTGTCGCGCAAGACTTGAAGAAAGCACTTGGTCTTCCAGTATTTCTTTCTTT 3810
Db 3649 AGCTTTTGTGTCGCGCAAGAACTGAAGAAAGCACTTGGTCTTCCAGTATTTCTTTCTTT 3708

3811 TACACAGATGAGTCAGTGCCTGTTAGCACCTCCACAAAATTTGACCTTTAAATCAACTGTACT 3870
3709 TACACAGATGAGTCAGTGCCTGTTAGCACCTCCACAAAATTTGACCTTTAAATCAACTGTACT 3768
3871 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTTCCAGGTGGTATTTAAAGTA 3930
3769 TCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTTTCCAGGTGGTATTTAAAGTA 3828
3931 TATAGTTTTAAATTCATCAACATGAAATCTGCACCTATATATTTAAGAAATATATCAGGA 3990
3829 TATAGTTTTAAATTCATGAAACATGAACTGCACCTATATATTTAAGAAATATATCAGGA 3888
3991 TTTAAACTGAAGCCAAACTTTGTGACTGGAAACCAAGTCAGCACCTACTCTATCCGTGTA 4050
3889 TTTAAACTGAAGCCAAACTTTGTGACTGGAAACCAAGTCAGCACCTACTCTATCCGTGTA 3948
4051 TCTGCCTTCACAAAGTTGGAATGGCAATCAATTTAGTAATGTAGTAATAAATTCACAACC 4110
3949 TCTGCCTTCACAAAGTTGGAATGGCAATCAATTTAGTAATGTAGTAATAAATTCACAACC 4008
4111 CAAGAAATCAGTCCAGATCGTGCAGAAATATGCAAGTCATGGCACTAGCTGGCAGTCA 4170
4009 CAAGAAATCAGTCCAGATCGTGCAGAAATATGCAAGTCATGGCACTAGCTGGCAGTCA 4068
4171 GTTTTGTAGTGAATGGATCCACCCAAAAGGCAATGGAATTAATACGCAAGTATATGTA 4230
4069 GTTTTGTAGTGAATGGATCCACCCAAAAGGCAATGGAATTAATACGCAAGTATATGTA 4128
4231 ACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGATCAATGTACACTTTTCATAAAG 4290
4129 ACAGTTGAAAGGAATTTCTACAAAAGTTTCTCCCAAGATCAATGTACACTTTTCATAAAG 4188
4291 CTTCTTGCCAATACCTCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGT 4350
4189 CTTCTTGCCAATACCTCATATGCTTTTAAAGTAAGAGCTTCAACCTCAGCTGGTGAAGT 4248
4351 GATGAAAGCACATGCCATGTCCAGCACATACCTTGAAAACAGTTCCCAAGTGTCCCAAAAT 4410
4249 GATGAAAGCACATGCCATGTCCAGCACATACCTTGAAAACAGTTCCCAAGTGTCCCAAAAT 4308
4411 ATTGCTTTTCTGATGTTTCAAGTCAACTAGTGCAACATTTGCAATGGATAAGACTGACACT 4470
4309 ATTGCTTTTCTGATGTTTCAAGTCAACTAGTGCAACATTTGCAATGGATAAGACTGACACT 4368
4471 ATCTCTGGCTACTTTTCAAAAATTTACAAAATTTACCACTCAACTCTCGTCTCABAAAATGCAAA 4530
4369 ATCTCTGGCTACTTTTCAAAAATTTACAAAATTTACCACTCAACTCTCGTCTCABAAAATGCAAA 4428
4531 GAATGGGAATCCGAAGAATGTGTTGAATATCAAAAATTTCAATACCTCTATGAAGCTCAC 4590
4429 GAATGGGAATCCGAAGAATGTGTTGAATATCAAAAATTTCAATACCTCTATGAAGCTCAC 4488
4591 TTAACCTGAAGAGACAGTATATGGATTTAAAGAAATTTAGATGGTATAGATTTCAAGTGGCT 4650
4489 TTAACCTGAAGAGACAGTATATGGATTTAAAGAAATTTAGATGGTATAGATTTCAAGTGGCT 4548
4651 GCCAGACCAATGCTGGCTATGGCAATGCTTCAACTGCAATTTCTACAAAATCTGCTCT 4710
4549 GCCAGACCAATGCTGGCTATGGCAATGCTTCAACTGCAATTTCTACAAAATCTGCTCT 4608
4711 GGCCCTCCAGATGGTCTCTCGAAAATGTTTCATGTAGTAGCAACATCACCTTTTAGCATC 4770
4609 GGCCCTCCAGATGGTCTCTCGAAAATGTTTCATGTAGTAGCAACATCACCTTTTAGCATC 4668
4771 AGCATAACTGGAGTGAACCTGCTGTCTATTTACTGGACCAACATGTTTATCTGATGATGTC 4830
4669 AGCATAACTGGAGTGAACCTGCTGTCTATTTACTGGACCAACATGTTTATCTGATGATGTC 4728
4831 AAATCGGTAGATAATGATGAAATTTATATATATCTCTCATCAAGTCAAAATGAAGAAATATAA 4890
4729 AAATCGGTAGATAATGATGAAATTTATATATATCTCTCATCAAGTCAAAATGAAGAAATATAA 4788
4891 ACCATAGAAATTTAAAGATTTAGAAAATATTTCACAAAGGTAATCTGTAGTATCATCTGCATTT 4950

4789 ACCATAGAAATTTAAAGATTTAGAAAATTTCAAGGTATTTCTGTAGTATCATCTGCATTT 4848
4951 ACTGGGAACATTTAGTCTGCTGCATATCTAGAAAGGAAGTCAAGTCTGGAATGATTTGTTACT 5010
4849 ACTGGGAACATTTAGTCTGCTGCATATCTAGAAAGGAAGTCAAGTCTGGAATGATTTGTTACT 4908
5011 ACTTTAGAAATCAGCCCCAAAAGGACCCACCTAACCAACATGACATTTTCAGAAAGTACAGAT 5070
4909 ACTTTAGAAATCAGCCCCAAAAGGACCCACCTAACCAACATGACATTTTCAGAAAGTACAGAT 4968
5071 GAAGTTTACAAAATTTCAATTAACAGTTCCTTCTCTCTCTCAACCTTAATGGAAATATCCAA 5130
4969 GAAGTTTACAAAATTTCAATTAACAGTTCCTTCTCTCTCTCAACCTTAATGGAAATATCCAA 5028
5131 GTATATCAAGCTCTGTTTACCGAGAAGTATCTTACTGCTGCTCCAGATTCACAACTTC 5190
5029 GTATATCAAGCTCTGTTTACCGAGAAGTATCTTACTGCTGCTCCAGATTCACAACTTC 5088
5191 AGTATTTATACAGAAAACCAACACATTCGTTCATTGCAATGCTAGAAAGGACTAAAAGTGA 5250
5089 AGTATTTATACAGAAAACCAACACATTCGTTCATTGCAATGCTAGAAAGGACTAAAAGTGA 5148
5251 CATACATACAATATCAGTGTTTTACGAGTCAATAGTGTGCTGAGTCCAAAAGTTCGG 5310
5149 CATACATACAATATCAGTGTTTTACGAGTCAATAGTGTGCTGAGTCCAAAAGTTCGG 5208
5311 ATGAGAAATACCATGATATCAAGCTCCAGCACGACCAACCAACCAACCAACCCCTATT 5370
5209 ATGAGAAATACCATGATATCAAGCTCCAGCACGACCAACCAACCAACCAACCCCTATT 5268
5371 TATGATGCCACAGAAAACCTGCTGTGACTTCAACAAATTAACAATCAGAAATGCAATA 5430
5269 TATGATGCCACAGAAAACCTGCTGTGACTTCAACAAATTAACAATCAGAAATGCAATA 5328
5431 TGTTACTACAGTATGATCATGGACCAATATAAATAATGTACAAGTCTTGGCAGAGAAACA 5490
5329 TGTTACTACAGTATGATCATGGACCAATATAAATAATGTACAAGTCTTGGCAGAGAAACA 5388
5491 GGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGATATTTTAATAAAGCAAGG 5550
5389 GGAGCTCAGCATGATGGAATGTAAACAAAGTGTATGATGATATTTTAATAAAGCAAGG 5448
5551 CCATATTTTACAAATGAAAGGCTTTCTTAAACCTCCATGTACAGAAAGGAAAGCAAAAGTTT 5610
5449 CCATATTTTACAAATGAAAGGCTTTCTTAAACCTCCATGTACAGAAAGGAAAGCAAAAGTTT 5508
5611 AGTGCAATGAAGAAATCTACATCATAGTGTCTGATTAATGCAATGATTCCTGGCAAT 5670
5509 AGTGCAATGAAGAAATCTACATCATAGTGTCTGATTAATGCAATGATTCCTGGCAAT 5568
5671 GAAGCAAAATTTGCAATGGACCACTGAAACCAAAAAGGCAATATCTTATTAATTTAGA 5730
5569 GAAGCAAAATTTGCAATGGACCACTGAAACCAAAAAGGCAATATCTTATTAATTTAGA 5628
5731 GCTACAAATATTTATGGGCAATTTTACTGACTCTGATTAATTTCTGACCTGTTTAAAGCTTTA 5790
5629 GCTACAAATATTTATGGGCAATTTTACTGACTCTGATTAATTTCTGACCTGTTTAAAGCTTTA 5688
5791 GGGGAAGGACTTTTACAGAAAGAAACCGTAGAGATCATTTCTTTCCGTCACTTTTGTATTCCTT 5850
5689 GGGGAAGGACTTTTACAGAAAGAAACCGTAGAGATCATTTCTTTCCGTCACTTTTGTATTCCTT 5748
5851 TCAATAATTTCTTTGGAAACAGCTATTTTGGCAATTTGCAAGAAATTCGACAGAAAGCAGAA 5910
5749 TCAATAATTTCTTTGGAAACAGCTATTTTGGCAATTTGCAAGAAATTCGACAGAAAGCAGAA 5808
5911 GAAGGTGGCACATCTCTCTCAGGATGACAGAAATTTTGTGACATTAATTTGAAGTGGAT 5970
5809 GAAGGTGGCACATCTCTCTCAGGATGACAGAAATTTTGTGACATTAATTTGAAGTGGAT 5868
5971 CAGCTCATCAGAGTGGCAGACCTGGAACTGGAAGGACGAGAGATTTAAACCGG 6020

QY 316 GGGATCTCTGCTTGGGAATACACCACCTAAATCCAAATGGGAAGATTATATCTTACATT 375
DB 235 GGGATCTCTGCTTGGGAATACACCACCTAAATCCAAATGGGAAGATTATATCTTACATT 294
QY 376 GTCAATATPAAGGAAGTTGTCGGTGGATGCAACACAGTATATACACAAGTCAGATCAAAAG 435
DB 295 GTCAATATPAAGGAAGTTGTCGGTGGATGCAACACAGTATATACACAAGTCAGATCAAAAG 354
QY 436 CAGACAGTCGGAAGTTCTTCTTACTAATCTTAATCTTAAATCCCTGGGAACAACATATGAATTAAG 495
DB 355 CCAGACAGTCGGAAGTTCTTCTTACTAATCTTAAATCCCTGGGAACAACATATGAATTAAG 414
QY 496 GTTGTGCTGTAACACAGTCGTCGCAATGGAGTGTGTTAGTGATTCATTTCTCTTCCAAACT 555
DB 415 GTAGCTGCTGTAACACAGTCGTCGCAATGGAGTGTGTTAGTGATTCATTTCTTCCAAACT 474
QY 556 GCAGAAAGTGCTCCAGGAAAGTGGTGAATCTCACAGTTTGAGGCTTACAAAGCTTCAGCA 615
DB 475 GCAGAAAGTGCTCCAGGAAAGTGGTGGATTTTCACAGGTGAGGCTGTC--CCGTTTCAGCA 532
QY 616 GTTAAGCTGATTTGGTATTTAATCTCGGCAACCAATGGCAAAATACAGAGCTTCAAGATT 675
DB 533 G-TAAGCTGATGGTA--TAACCTCGGCAACCAAA-AAAAAAATTAACCAAGCTTCAAGATT 588
QY 676 AGTGTCAAAACATGCCAGAAAGTGGATAGTAGTGAAAGATGTTCTCAATCAGAGTAGAGGAC 735
DB 589 AGTGTCAAGCATACAGAAAGTGGATAGTAGTGAAAGATGTTCAATCAGAGTAGAGTGC 648
QY 736 ATTTTGACTGGGAAATGGC---AGAAATGCAATGGAATAGTGAATCTTTTTTATGGAGT 792
DB 649 ATTTTAAGTGCTTCCCTTCTTGGCACTGCAACGAGAAATAGTGAATCTTTTTTATGGAGT 708
QY 793 ACAGCCAGCCCTTCTCCAAACCTTGGTAGAGTTA GACCTCCATCGGCTACCAACATTCAC 852
DB 709 ACAGCCAGCCCTTCTCCAAACCTTGGTAGAGTTTACACCTCCATCGGCTACCAACATTCAC 768
QY 853 TCAAGCAGCTTGACACAGAAATGAGATCAGCTCTGTGTGGAAGAGCTATCAGTTTGTGA 912
DB 769 TCAGCAGCTTGACACAGAAATGAGATCAGCTCTGT---GAAGAGCTATCAGTTTGTGA 825
QY 913 GTGACACACTTGAGACCTTTATACAACATATCTTTTTTGAAGTTTCAGCTGCTACAACTGAA 972
DB 826 GTGACACACTTGAGACCTTTATACAACATATCTTTTTTGAAGTTTCAGCTGCTACAACTGAA 885
QY 973 GCAGGTTATTTGATAGTAGATTGTACAGAACACCAAGATCAGTGCCTGGAAGCACCA 1032
DB 886 GCAGGTTATTTGATAGTAGATTGTACAGAACACCAAGATCAGTGCCTGGAAGCACCA 945
QY 1033 CAAAACTGGGTACAGCAACATCACAGGAAGTCTCTTTCAATTTTATGGGACCCACCA 1092
DB 946 CAAAATCGGTAAACAGCAACATCACAGGAAGTCTCTTTCAATTTTATGGGACCCACCA 1005
QY 1093 ACTATAGTAACAGGAAATTTAGTTATAGAGTTGAATTTATATGGAACCAT---CAGGTGCG 1149
DB 1006 ACTATAGTAACAGGAAATTTAGTTATAGAGTTGAATTTATATGGAACCATCAGCAGGTGCG 1065
QY 1150 ATTTTGGATTAACAGCAACAAAGACCTCAAGTTTGCATTTCACTAACTTAACACATTTTACA 1209
DB 1066 ATTTTGGATTAACAGCAACAAAGACCTCAAGTTTGCATTTCACTAACTTAACACCAATTTTACA 1125
QY 1210 ATGTATAGTCTATATTGGGCTGAAACACAGTGCAAGGACCTGGGGCCCAAGTCAAAATATT 1269
DB 1126 ATGTATAGTCTATATTGGGCTGAAACACAGTGCAAGGACCTGGGGCCCAAGTCAAAATATT 1185
QY 1270 TCAGTATTCCTCCACACAGATGTTCCAGGGGCAAGTGTGTTGATTTACAACTTCGAGAGGTA 1329
DB 1186 TCAGTATTCCTCCACACAGATGTTCCAGGGGCAAGTGTGTTGATTTTACAACCTTGAATTTTAAAC 1245
QY 1330 GAATCCACGCAAGTAAAGATTACTTGGAGAAACCAACGACCAACCAATGGAATTTAATTAAC 1389
DB 1246 GAATCCACGCAAGTAAAGATTACTTGGAGAAACCAACGACCAACCAATGGAATTTAATTAAC 1305

QY 1390 CAATACCGAGTGAAGTGTCTAGTTCCAGAGACAGGAATAATTTTGGAAAAATCTTTGTCTC 1449
DB 1306 CAATACCGAGTGAAGTGTCTAGTTCCAGAGACAGGAATAATTTTGGAAAAATCTTTGTCTC 1365
QY 1450 ACTGGAATAATAGTATATAAATGACCCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1509
DB 1366 ACTGGAATAATAGTGA---ATAAATGACCCCATGGCTCCAGAAATTTGTGAACATAGTAGAG 1422
QY 1510 CCAATGGTAGGATTATATAGGGTTTCAGCAGAGATGTCGTCTGACCTTCACCTCACTTGCT 1569
DB 1423 CCAATGGTAGGATTATATAGGGTTTCAGCAGAGATGTCGTCTGACCTTCACCTCACTTGCT 1482
QY 1570 ACATTTATATAPAACAGCCATCCAGATAAAAATCTTTCTGCAAGGAATAGAGCTGAAGAC 1629
DB 1483 ACATTTATATAPAACAGCCATCCAGATAAAAATCTTTCTGCAAGGAATAGAGCTGAAGAC 1542
QY 1630 CAGACTTCACCGTTGTAACTCAAGGAATCAGTATATTTACTGACATTTGCAGCTGAACAG 1689
DB 1543 CAGACTTCACCGTTGTAACTCAAGGAATCAGTATATTTACTGACATTTGCAGCTGAACAG 1602
QY 1690 CTGCTCTTATGTTATCAGGAGACTTGTACTTTTCACTGAGCACATGATTTAGTGTATCTGCT 1749
DB 1603 CTGACTTATGTTCTTATCAGATTAAGGAGATTTTGGGCTGACACAAATGGGGTTTCTAGA 1662
QY 1750 TTCCACATCATGGGAGAGGACCCACCAACAGTTCTCAGTGTGTAGGACACGTCAGCAAGTG 1809
DB 1663 TATACAAATCATG-----TCATCTGCAAGCAGGGACAAATTCAGCTTTCCCGCAGGC 1710
QY 1810 CCAAGCTCCATTAATAATTTATAAACTATAAATAATATTAGTTCTTCTCATTTATTTTGTATAT 1869
DB 1711 CCTTTGTGAGCCCAAAATTTTCAGAGTTTACACATGTTTACNTAACAGAGATATTTTACAC 1770
QY 1870 TGGGATCCTCCAGATATATCCCAATGGAATAATACTCATATACGATTTATGCAATGGAA 1929
DB 1771 TGGGATCCTCCAGATCCTGTATTTTTCATCATTTACCTTTATCACATTTTGTGATGTA 1830
QY 1930 TTGGATCAAAAACAGAGCATTCCAGATAACTACCATAGATAACAGAGTTTCTCATACAGGG 1989
DB 1831 AACCAATCCAAGATTTATTTTAAAGGACATTTAAACAGTTTGTCTCTTCTTATAGGG 1890
QY 1990 TTAAAGAAATACACAAAATACAAATAGAGTGGCAGCTCAACCCACGATGGAGAAAGT 2049
DB 1891 TTAAAGAAATACACAAAATACAAATAGAGTGGCAGCTCAACCCACGTTGGAGAAAGT 1950
QY 2050 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAGAGATGAACCCGAAATCATCA 2109
DB 1951 TCTTTGTCTGAAGAAATGACATCTTTGTGAGAACTTCAGAGAGATGAACCCGAAATCATCA 2010
QY 2110 CCTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAATAAGGTTGAAGTGGTCA 2169
DB 2011 CCTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAATAAGGTTGAAGTGGTCA 2070
QY 2170 CCGGAAAGCCCAATGGGATCATTATTTGCTTATGAAGTGTCTATATAAAAAATATAGATCT 2229
DB 2071 CCGGAAAGCCCAATGGGATCATTATTTGCTTATGAAGTGTCTATATAAAAAATATAGATCT 2130
QY 2230 TTATATATGAAGAACACATCAACACAGACATATATTAAGGAACTTAAGACCTCACACC 2289
DB 2131 TTATATATGAAGAACACATCAACACAGACATATATTAAGGAACTTAAGACCTCACACC 2190
QY 2290 CTCTATAACATTTCTGTAAGGTTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT 2349
DB 2191 CTCTATAACATTTCTGTAAGGTTTACACAGATTTGGTCAATGGCAATCAGGTATCTTCT 2250
QY 2350 TTACTCTCTGTAAGGACTTCGGAGATGTCCTGTATAGTGCACCAAGAAATATCACTTAC 2409
DB 2251 TTACTCTCTGTAAGGACTTCGGAGATGTCCTGTATAGTGCACCAAGAAATATCACTTAC 2310
QY 2410 AAAAAATTTCTTCTGAGAGATTCAGCTATCATCTTCTTCCCGCAAGTAGTCCCAATGGA 2469
DB 2311 AAAAAATTTCTTCTGAGAGATTCAGCTATCATCTTCTTCCCGCAAGTAGTCCCAATGGA 2370
QY 2470 ATCATAAAAAATATATACAAATTTTCTCAAGAGAGATTAATGGAAATGAGGAAAGAACTATA 2529

Db 2371 ATCATCAAAAAATATACAAATTTATCTCAAGAGAGTAATGGAAATGAGAAAGAACTATA 2430
QY AATACAACTCTTTTAAACCAAAACATTTAAAGTACTGAAGAAATATACCCAAATATATCAT 2589
Db 2431 AATACAACTCTTTTAAACCAAAACATTTAAAGTCTGAAGAAATATACCCAAATATATCAT 2490
QY GAGGTCTCTGCTAGTACATGCTAAAGGTGAAGAGTTCGAGTCTCCCAATATATATATG 2649
Db 2491 GAGGTCTCTGCTAGTACATGCTAAAGGTGAAGAGTTCGAGTCTCCCAATATATATG 2550
QY ACGGAGGAGAGTCTCTGATTTCTCCCTCTCAAGACTTCTCTGTAAGACAGTTGCTGCT 2709
Db 2551 ACGGAGGAGAGTCTCTGATTTCTCCCTCTCAAGACTTCTCTGTAAGACAGTTGCTGCT 2610
QY GTCCAGGCTGAAGTTGTCATGGCAACCAACCCCTGGAGCCAAATGGAAATATCTTTATTTAC 2769
Db 2611 GTCCAGGCTGAAGTTGTCATGGCAACCAACCCCTGGAGCCAAATGGAAATATCTTTATTTAC 2670
QY ACAGTTTATGTCTGGAAATAGATCATCATTTAAATAATTTAATGTCTCACTGAAACATCATTTG 2829
Db 2671 ACAGTTTATGTCTGG --- AGATCATCATTTAAATAATTTAATGTCTCACTGAAACATCATTTG 2727
QY GAGTTATCAGATTGGAATTAATGTTGAATACAGTGTCTTATGTAAACAGTCTAGCACCCAGA 2889
Db 2728 GAGTTATCAGATTGGAATTAATGTTGAATACAGTGTCTTATGTAAACAGTCTAGCACCCAGA 2787
QY TTTGGTGATGGGAAACAGAGCAATATCATTTAGCTTTTCAAAACACAGAGGAGCACCA 2949
Db 2788 TTTGGTGATGGGAAACAGAGCAATATCATTTAGCTTTTCAAAACACAGAGG --- ACCA 2844
QY AGCGATCCCTCCCAAGATGTTTATATGCAAACTCAGTTCTTCAATCAATATCTTTTC 3009
Db 2845 AGCGATCCCTCCCAAGATGTTTATATGCAAACTCAGTTCTTCAATCAATATCTTTTC 2904
QY TGGACACTCTCTTCAAAACCTAATGGGATTTATCAATATTTACTCTGTTTATTTACAGAAAT 3069
Db 3010 TGGACACTCTCTTCAAAACCTAATGGGATTTATCAATATTTACTCTGTTTATTTACAGAAAT 2964
QY ACTTCAGGTACTTTTATGAGAAATTTTACACTCTCAATGAATTAACCAATGACTTTGACAAAT 3129
Db 2965 ACTTCAGGTACTTTTATGAGAAATTTTACACTCTCAATGAATTAACCAATGACTTTGACAAAT 3024
QY ATGACTGTATCCCAATATAGATAAACTGACAAATTTAGCTTACTATATACATTTTGGTTA 3189
Db 3025 ATGACTGTATCCCAATATAGATAAACTGACAAATTTAGCTTACTATATACATTTTGGTTA 3084
QY ACAGCAAGTACTTTCAGTTGGAAATGGGAAATAAAGCAGTGACATCAATGAAATATACACA 3144
Db 3190 ACAGCAAGTACTTTCAGTTGGAAATGGGAAATAAAGCAGTGACATCAATGAAATATACACA 3249
QY GATCAAGACATACCTGAAGGTTTGTGGAAACCTGACATTCAGAAATCCATTTCTGCAACT 3309
Db 3145 GATCAAGACATACCTGAAGGTTTGTGGAAACCTGACATTCAGAAATCCATTTCTGCAACT 3204
QY GCAATAAATGTAAGTGGGTCACCGGCTCAACCAACCGGCTAGTCTTCTACTATGTT 3369
Db 3205 GCAATAAATGTAAGTGGGTCACCGGCTCAACCAACCGGCTAGTCTTCTACTATGTT 3264
QY TCACTGATCTTACAGCAGACTCTCTGCAATGTGAGACCACTCTTGTATACATATGAGAGA 3429
Db 3265 TCACTGATCTTACAGCAGACTCTCTGCAATGTGAGACCACTCTTGTATACATATGAGAGA 3324
QY AGCATATATTTGATATCTGGAAATACACTGATTTATATATTAATAAATTAATCTCCATCA 3489
Db 3325 AGCATATATTTGATATCTGGAAATACACTGATTTATATATTAATAAATTAATCTCCATCA 3384
QY ACAGAAAGGGATTTCTGATACCTATATGCCCCAGCTATACATCAAGACTGAAGAAGAT 3549
Db 3385 ACAGAAAGGGATTTCTGATACCTATATGCCCCAGCTATACATCAAGACTGAAGAAGAT 3444
QY GTCCCCAGAACTTCAACCAATATCAACATTTTAAATAAATCTTTCTCTACTCCTCAGTTCTC 3609

Db 3445 ATCCGAGAACTTTCACCAATATCAACACTTTTAAATAAATCTTCTCTACTCCTCAGTTCTC 3504
QY TTATCATGGGATCCCCAGTAAAGCCAAATAGTGTCAATAAATAGTATATGATTTAACTTTA 3669
Db 3505 TTATCATGGGATCCCCAGTAAAGCCAAATAGTGTCAATAAATAGTATATGATTTAACTTTA 3564
QY CHAGGACCAATGAAATATTTCTTTTCAATCTCTGATTAATTAATATATATTTGGAAGAG 3729
Db 3565 CHAGGACCAATGAAATATTTCTTTTCAATCTCTGATTAATTAATATATATTTGGAAGAG 3624
QY CTTTCCAGTATTTACATATATAGCTTTTGTCTGCGCAAGAACTAGAAAAAGGACTTGTGT 3789
Db 3625 CTTTCCAGTATTTACATATATAGCTTTTGTCTGCGCAAGAACTAGAAAAAGGACTTGTGT 3684
QY CTTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCAAGTGTGAGTGTGAGTGTGAGTGTG 3849
Db 3685 CTTTCCAGTATTTCTTTTCTTTTACACAGATGAGTCAAGTGTGAGTGTGAGTGTGAGTGTG 3744
QY TTGACTTTAATCAACTGTACTTTCAGACTTTGTATGGCTGAAATGAGGCCAAGTCTCTCTT 3909
Db 3745 TTGACTTTAATCAACTGTACTTTCAGACTTTGTATGGCTGAAATGAGGCCAAGTCTCTCTT 3804
QY CCAGGTGGTATTTAAAGTATATAGTTTAAATTTCAATGAACATGAACATGAACATGAACATGA 3969
Db 3805 CCAGGTGGTATTTAAAGTATATAGTTTAAATTTCAATGAACATGAACATGAACATGAACATGA 3864
QY TATTAATAAGAAATATACAGGATTTAAACTGAAAGCCAAACTTTGTTGGACTGGAACCAAGTC 4029
Db 3865 TATTAATAAGAAATATACAGGATTTAAACTGAAAGCCAAACTTTGTTGGACTGGAACCAAGTC 3924
QY AGCACTTACTTATCCGTGTATCTCGGTTTACCAAAAGTTGGAATGGAATGGAATGGAATGGAAT 4089
Db 3925 AGCACTTACTTATCCGTGTATCTCGGTTTACCAAAAGTTGGAATGGAATGGAATGGAATGGAAT 3984
QY AATGTAGTAAATTTCAACCAACCAAGATTCAGTTTCCAGATGTCGTGAGAAATATGCAAGTGC 4149
Db 3985 AATGTAGTAAATTTCAACCAACCAAGATTCAGTTTCCAGATGTCGTGAGAAATATGCAAGTGC 4044
QY ATGGCAACTAGCTGGCAGTCAAGTTTGTAGTAAATGGGATCCACCCAAAGGCAAAATGGA 4209
Db 4045 ATGGCAACTAGCTGGCAGTCAAGTTTGTAGTAAATGGGATCCACCCAAAGGCAAAATGGA 4104
QY ATAATAAGCAGTATATGGTAAAGGAAATCTAGAAAGTTTCTCCCAAGAT 4269
Db 4105 ATAATAAGCAGTATATGGTAAAGGAAATCTAGAAAGTTTCTCCCAAGAT 4164
QY CACATGTACACTTTCATAAAGCTTCTGCCAATCTCTCATATGTTCTTTAAAGTAAAGCT 4329
Db 4165 CACATGTACACTTTCATAAAGCTTCTGCCAATCTCTCATATGTTCTTTAAAGTAAAGCT 4224
QY TCAACCTCAGCTGGTGAAGGATGAAAGCAATGCCATGTGAGCACTACTACTGAAACA 4389
Db 4225 TCAACCTCAGCTGGTGAAGGATGAAAGCAATGCCATGTGAGCACTACTACTGAAACA 4284
QY GTTCCAGTGTCCCAAAATATTTCTTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTT 4449
Db 4285 GTTCCAGTGTTCACCAAAATATTTCTTTTCTGATGTTTCTGATGTTTCTGATGTTTCTGATGTTT 4344
QY ACATGGAATGAGCTGTGACACTATCTCTGGCTACTTTTCAAAATTAATAAATTAATAAATTAATA 4509
Db 4345 ACATGGAATGAGCTGTGACACTATCTCTGGCTACTTTTCAAAATTAATAAATTAATAAATTAATA 4404
QY CTTTCTGTCTCAAAATGCAAGAAATGGGAAATCCGAAGAAATGTTTGAATATCAAAAAATTT 4569
Db 4405 CTTTCTGTCTCAAAATGCAAGAAATGGGAAATCCGAAGAAATGTTTGAATATCAAAAAATTT 4464
QY CAATACCTCTATGAAAGCTCACTTAATCTGAAAGAGA CAGTATATGGAATTAAGAAATTTAGA 4629
Db 4465 CAATACCTCTATGAAAGCTCACTTAATCTGAAAGAGA CAGTATATGGAATTAAGAAATTTAGA 4524
QY TGGTATAGTTCCAAAGTGGCTGCCAGCACCAATGCTGGCTATGGAATGGAATGGAATGGAATGGA 4689
Db 4525 TGGTATAGTTCCAAAGTGGCTGCCAGCACCAATGCTGGCTATGGAATGGAATGGAATGGAATGGA 4584

QY	4690	ATTTCCTACAAACCTGCTGCGCCCTCCAGATGGTCTCCTGAAATGTTTCATGTAGTA	4749
DB	4585	ATTTCCTACAAACCTGCTGCGCCCTCCAGATGGTCTCCTGAAATGTTTCATGTAGTA	4644
QY	4750	GCAACATCACCTTTTAGCATCAGCATAAAGCTGAGTGAACCTGCTGTCATTAATCTGGACCA	4809
DB	4645	GCAACATCACCTTTTAGCATCAGCATAAAGCTGAGTGAACCTGCTGTCATTAATCTGGACCA	4704
QY	4810	ACATGTTATCTGATGATGTCATAATCGGTGAGTAAATGATGAATTAATATATCTTCATC	4869
DB	4705	ACATGTTATCTGATGATGTCATAATCGGTGAGTAAATGATGAATTAATATATCTTCATC	4764
QY	4870	AACTCAAAATGAAGAAATAAACCATAGAAATTAAGATTTAGAAATATTTCAAGGTTAT	4929
DB	4765	AACTCAAAATGAAGAAATAAACCATAGAAATTAAGATTTAGAAATATTTCAAGGTTAT	4824
QY	4930	TCTGTAGTCATCACTGCAATTTACTGGGAACATTTAGTGTGTCATATGTAGAGGGAATCA	4989
DB	4825	TCTGTAGTCATCACTGCAATTTACTGGGAACATTTAGTGTGTCATATGTAGAGGGAATCA	4884
QY	4990	AGTGCTGAATGATGTTACTACTTTAGAAATCAGCCCCAAAGGACCCACTAAACAATG	5049
DB	4885	AGTGCTGAATGATGTTACTACTTTAGAAATCAGCCCCAAAGGACCCACTAAACAATG	4944
QY	5050	ACATTTCAAGATACCCAGATGAAGTTACAAAATTTCAATTTAAGCTTCTCTCTCTCT	5109
DB	4945	ACATTTCAAGATACCCAGATGAAGTTACAAAATTTCAATTTAAGCTTCTCTCTCTCTCT	5004
QY	5110	CAACCTAAATGAAATATCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTACT	5169
DB	5005	CAACCTAAATGAAATATCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTACT	5064
QY	5170	GCTGTCAGATTCACAACTCAGTATATACAGAAAACCAACACATTTGTCATTCGAATG	5229
DB	5065	GCTGTCAGATTCACAACTCAGTATATACAGAAAACCAACACATTTGTCATTCGAATG	5124
QY	5230	CTAGAAGGACTAAAGGTGGACATACATACATATCAGTGTTCAGCAGTCAATAGTGT	5289
DB	5125	CTAGAAGGACTAAAGGTGGACATACATACATATCAGTGTTCAGCAGTCAATAGTGT	5184
QY	5290	GGTGAGGTCCTAAAGTTCAGATGAGATTAACATGATATCAAGCTCCAGCAGACCA	5349
DB	5185	GGTGAGGTCCTAAAGTTCAGATGAGATTAACATGATATCAAGCTCCAGCAGACCA	5244
QY	5350	AAAAACCAACCCCTATTTATGATGCCACAGGAAAACCTGCTGTGACTTTCAACAACA	5409
DB	5245	AAAAACCAACCCCTATTTATGATGCCACAGGAAAACCTGCTGTGACTTTCAACAACA	5304
QY	5410	ATTACATCAGATGCCAATATGTTACTACAGTGTATGATGAGACCAATTAATAATGTA	5469
DB	5305	ATTACATCAGATGCCAATATGTTACTACAGTGTATGATGAGACCAATTAATAATGTA	5364
QY	5470	CAAGTCTTGGACAGAAACAGGAGCTCAGCATGATGGAATGTAACAAAGTGTATGAT	5529
DB	5365	CAAGTCTTGGACAGAAACAGGAGCTCAGCATGATGGAATGTAACAAAGTGTATGAT	5424
QY	5530	GCATATTTTAAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCCATGT	5589
DB	5425	GCATATTTTAAAGCAAGGCCATATTTTACAAATGAAGGCTTTCTTAACCCCTCCATGT	5484
QY	5590	ACAGAAGGAACAGAAAGTTTAGTGGCAATGAAGAAATCTACATATAGTGTCTGATAT	5649
DB	5485	ACAGAAGGAACAGAAAGTTTAGTGGCAATGAAGAAATCTACATATAGTGTCTGATAT	5544
QY	5650	GCATGATCATTTCTGGCAATGAAGCAAAAATTTGCAATGGACCACTGAAACCAAAAAG	5709
DB	5545	GCATGATCATTTCTGGCAATGAAGCAAAAATTTGCAATGGACCACTGAAACCAAAAAG	5604
QY	5710	CAATACTTTATTTAAATTTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATAT	5769
DB	5605	CAATACTTTATTTAAATTTTAGAGCTACAAATATTTATGGGCAATTTTACTGACTCTGATAT	5664

QY	5770	TCTGACCCCTGTTTAAAGCTTTTAGGGGAAGGACCTTTTTCAGAAAGAACCGTAGAGATCATTTCT	5829
DB	5665	TCTGACCCCTGTTTAAAGCTTTTAGGGGAAGGACCTTTTTCAGAAAGAACCGTAGAGATCATTTCT	5724
QY	5830	TCCGTCACTTTGTGTATCTCTTCAATAATTTCTCTTGGAAACAGCTATTTTGTCAATTTGCA	5889
DB	5725	TCCGTCACTTTGTGTATCTCTTCAATAATTTCTCTTGGAAACAGCTATTTTGTCAATTTGCA	5784
QY	5890	AGAAATTCAGACAGACAGAAAGAGGTGGCACTATCTCTCTCAGGATGACAGAAATTTAT	5949
DB	5785	AGAAATTCAGACAGACAGAAAGAGGTGGCACTATCTCTCTCAGGATGACAGAAATTTAT	5844
QY	5950	GACACTAAATTTGAAGCTGATCAGCTCATCAGTGGCAGACCTCGAACTGAAGACGAG	6009
DB	5845	GACACTAAATTTGAAGCTGATCAGCTCATCAGTGGCAGACCTCGAACTGAAGACGAG	5904
QY	6010	AGATTAAACCGGCCCAATTAAGCAAGAAATCTCTTCTGCAACATGTTTGAAGAGCTTTTGCA	6069
DB	5905	AGATTAAACCGGCCCAATTAAGCAAGAAATCTCTTCTGCAACATGTTTGAAGAGCTTTTGCA	5964
QY	6070	AACAAACCACTTAAGTTTCAAGAGAAATTTTCGGAATTAACAAATTTTCTTCAGGATCTT	6129
DB	5965	AACAAACCACTTAAGTTTCAAGAGAAATTTTCGGAATTAACAAATTTTCTTCAGGATCTT	6024
QY	6130	TCTTCAACTGATGCTGATCTGCTTTGGAAATAGAGCAAAAACCGTTTCCCAACATAAAA	6189
DB	6025	TCTTCAACTGATGCTGATCTGCTTTGGAAATAGAGCAAAAACCGTTTCCCAACATAAAA	6084
QY	6190	CCATATAATAATAAATTAACAGAGTAAGCTGATGCTGACGCTAGTGTTCAGGTTCCGAT	6249
DB	6085	CCATATAATAATAAATTAACAGAGTAAGCTGATGCTGACGCTAGTGTTCAGGTTCCGAT	6141
QY	6250	TATATAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTCGTAATCAA	6309
DB	6142	TATATAATGCCAGCTATATTTCTGGTTATTTATGTCAAATGAATTTATTCGTAATCAA	6201
QY	6310	GCTCAGCTACAGAGAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACCAAGGCAAAA	6369
DB	6202	GCTCAGCTACAGAGAACAGTTGGAGATTTTGGAGAAATGTTGGGAAACCAAGGCAAAA	6261
QY	6370	ACATTAGTAATGCTTAACACAGTGTTTTGAAGAAAGGACGGATCAGATGCCATCAGTATGG	6429
DB	6262	ACATTAGTAATGCTTAACACAGTGTTTTGAAGAAAGGACGGATCAGATGCCATCAGTATGG	6321
QY	6430	CCAGAGGACAAACAGCCAGTATCTGCTTTTGGAGATATAGTGTATACAAAGCTATGAG	6489
DB	6322	CCAGAGGACAAACAGCCAGTATCTGCTTTTGGAGATATAGTGTATACAAAGCTATGAG	6381
QY	6490	GATGTTCAAAATAGATTGGACTATCAGGGATCTGAAATTTGAAGGATGCGGATGCAATG	6549
DB	6382	GATGTTCAAAATAGATTGGACTATCAGGGATCTGAAATTTGAAGGATGCGGATGCAATG	6441
QY	6550	ACTGTTTGCAAGTGTAACTTTTCTGCTGGCCAGAGCATGCGGTTCTTGAGAACAGCGCC	6609
DB	6442	ACTGTTTGCAAGTGTAACTTTTCTGCTGGCCAGAGCATGCGGTTCTTGAGAACAGCGCC	6501
QY	6610	CCTTAATTTCACTTTGTGAAGTTGGTTCGACAGACAGGACACATGACACACCTATG	6669
DB	6502	CCTTAATTTCACTTTGTGAAGTTGGTTCGACAGACAGGACACATGACACACCTATG	6561
QY	6670	ATTGTTCACTGCGAGTGGAGATTTGGAGAACTGGAGTTTTTATGCTCTGGACCATTTA	6729
DB	6562	ATTGTTCACTGCGAGTGGAGATTTGGAGAACTGGAGTTTTTATGCTCTGGACCATTTA	6621
QY	6730	ACACAAACATATAAATGACCATGATTTTGTGGATATATATGAGTATAGTGTGAACTGAGA	6789
DB	6622	ACACAAACATATAAATGACCATGATTTTGTGGATATATATGAGTATAGTGTGAACTGAGA	6681
QY	6790	AGTGAAGAAATGTCATGCTGGCAGAACTCTGGCAGTATATCTTTTACACAGTGCAT	6849
DB	6682	AGTGAAGAAATGTCATGCTGGCAGAACTCTGGCAGTATATCTTTTACACAGTGCAT	6741
QY	6850	CTGGATCTCTTATCAATTAAGGAAGTAATCAGCCCATCTGTTTTTGTAACTATTTCAGCA	6909

Db 6742 CTGATCTCTATCAATTAAGGAGTAATCAGCCCATCTGTTTGTAACTATTACGCA 6801
Qy 6910 CTTTCAAGATGACTCTTTTGACGCCATGGAA---GGTGATGTTGAGCTTGAATGGAA 6966
Db 6802 CTTTCAAGATGACTCTTTTGACGCCATGGAAAGGTGTCATGTTGAGCTTGAATGGAA 6861
Qy 6967 GAAACCACTATGTAATATTTCAGACCAAGATACAAATGGAAGAGNTTTTAAATCCCA 7026
Db 6862 GAAACCACTATGTAATATTTCAGACCAAGATACAAATGGAAGAGNTTTTAAATCCCA 6921
Qy 7027 GGGGCCAAAGTACCCCTCATCTCTCCGAATGAAATGCAACCTTAAAGAAATATCT 7086
Db 6922 GGGGCCAAAGTACCCCTCATCTCTCCGAATGAAATGCAACCTTAAAGAAATATCT 6981
Qy 7087 ATGCTTCTCTCAC 7099
Db 6982 ATGCTTCTCTCAC 6994

RESULT 7

US-10-218-779-5

; Sequence 5, Application US/10218779

; Publication No. US20040029222A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit

; APPLICANT: MacDougall, John

; APPLICANT: Millet, Isabelle

; APPLICANT: Ellerman, Karen

; APPLICANT: Stone, David

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William

; APPLICANT: Alsobrook II, John

; APPLICANT: Lepley, Denise

; APPLICANT: Rieger, Daniel

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Spytek, Kimberly

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Mishra, Vishnu

; APPLICANT: Fatturajan, Meera

; APPLICANT: Shenoy, Suresh

; APPLICANT: Rastelli, Luca

; APPLICANT: Tchernev, Velizar

; APPLICANT: Vernet, Corine

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Malyankar, Uriel

; APPLICANT: Guo, Xiaojia

; APPLICANT: Miller, Charles

; APPLICANT: Gangoli, Baha

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-214

; CURRENT APPLICATION NUMBER: US/10/218, 779

; CURRENT FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: 60/253,834

; PRIOR FILING DATE: 2000-11-29

; PRIOR APPLICATION NUMBER: 60/250, -926

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: 60/264,180

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/313,656

; PRIOR FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: 60/327,456

; PRIOR FILING DATE: 2001-10-05

; NUMBER OF SEQ ID NOS: 216

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 5

; LENGTH: 6994

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-218-779-5

Query Match 90.8%; Score 6453; DB 17; Length 6994;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 6770; Conservative 0; Mismatches 215; Indels 48; Gaps 14;

Qy 76 TGAATCTCTACTGGCTGAAATGTAATAAGATCGATTTCTTATCATTTTCTTTACTT 135
Db 1 TGAATCTCTACTGGCTGAAATGTAATAAGATCGATTTCTTATCATTTTCTTTACTT 60

Qy 136 TTTATTTGGGACTTTTCAGAGACACAGGTTTGAATGTTTCCAATGTCGTTCTCGTACTAGTAC 195
Db 61 TTTATTTGGGACTTTTCAGAGACACAGGTTTGAATGTTTCCAATGTCGTTCTCGTACTAGTAC 120

Qy 196 GATATAACCATCTCTTCAATTTCTACAAACATACCTTCACTGTTTACTAGATAGTAC 255
Db 121 GATATAACCATCTCTTCAATTTCTACAAACATACCTTCACTGTTTACTAGATAGTAC 180

Qy 256 CCAATGTAACAAACAGGCTTCCAGTCTTCTAGCCGGGGAAGAGTCGGATCTGCT 315
Db 181 TCTAATG-----AACAGGGCTTCCAGTCTTCTAGCCGGGGAAGAGTCGGATCTGCT 234

Qy 316 GGGATTTCTTCTGTTGGAAATACACCACTTAATCCAAATGGAAGGATTAATCTTACATT 375
Db 235 GGGATTTCTTCTGTTGGAAATACACCACTTAATCCAAATGGAAGGATTAATCTTACATT 294

Qy 376 GTCAAAATATAAGAAAGTTTGTCCGTGGATGCAAAACAGTATATATACAAAGTCAGATCA 435
Db 295 GTCAAAATATAAGAAAGTTTGTCCGTGGATGCAAAACAGTATATATACAAAGTCAGATCA 354

Qy 436 CCAGACAGTCTGGAGTTCTTCTTACTAATCTTAATCTTGGAAACATATGAATTAAG 495
Db 355 CCAGACAGTCTGGAGTTCTTCTTACTAATCTTAAATCTTGGAAACATATGAATTAAG 414

Qy 496 GTTGTCTGCTGAAACAGTCTGGCATTTGGAGTGTGTAGTATCATTTCTTCCAACT 555
Db 415 GTAGCTGCTGAAACAGTCTGGCATTTGGAGTGTGTAGTATCATTTCTTCCAACT 474

Qy 556 GCAGAAAGTCTCCAGGAAAGTGGTGAATCTCACAGTTTGAGGCTTCAACCGTTCAGCA 615
Db 475 GCAGAAAGTCTCCAGGAAAGTGGTGGATTTTCCAGAGTGAGGCTGTCT--CCGTTCAGCA 532

Qy 616 GTTAACTGATTTGGTATTTTAACTCCGCAACCAATGCGCAATTTACCAGTTCAGATT 675
Db 533 G-TAAGCTGATGTGTA--TACCTCGGCAACCA--AAAAAAATTTACCAGTTCAGATT 588

Qy 676 AGTGCAAAATGCTCCAGGAAAGTGGATAGTAGTGAAGAGTGTCTCAATCAGAGTAGAGGAC 735
Db 589 AGTGCAAAATGCTCCAGGAAAGTGGATAGTAGTGAAGAGTGTCTCAATCAGAGTAGAGGAC 648

Qy 736 ATTTTGAATGGGAAATTTGCC---AGAAATGCAATGAGATAGTGAATCTTTTTTATGGAGT 792
Db 649 ATTTTGAATGGTCTTCCCTTCTTGGCACTGCAACGAGATAGTGAATCTTTTTTATGGAGT 708

Qy 793 ACAGCCAGGCTTCTCCAACTTGTAGATAGTATACCTCCATCGGCTGACACATTTCA 852
Db 709 ACAGCCAGGCTTCTCCAACTTGTAGATAGTATACCTCCATCGGCTGACACATTTCA 768

Qy 853 TCAAGCAGCTTCACAGAGATGAGATCAGTCTGTGTGGAAGAGCTATCAGTTTGTGA 912
Db 769 TCAAGCAGCTTCACAGAGATGAGATCAGTCTGTGTGGAAGAGCTATCAGTTTGTGA 825

Qy 913 GTGACACACTTTGAGACTTTTATACAAATATCTTTTTTGAAGTTTTCAGTCTGCTCAACTGAA 972
Db 826 GTGACACACTTTGAGACTTTTATACAAATATCTTTTTTGAAGTTTTCAGTCTGCTCAACTGAA 885

Qy 973 GCAGGTTATATTGATAGTACGATTTGTGAGAACACAGATCAGTGTGCTGAGGACCA 1032
Db 886 GCAGGTTATATTGATAGTACGATTTGTGAGAACACAGATCAGTGTGCTGAGGACCA 945

Qy 1033 CAAAATGCTGTAACAGGCAACATCAGGAAAGTCTTTTCAATTTTATGGGACCA 1092
Db 946 CAAAATGCTGTAACAGGCAACATCAGGAAAGTCTTTTCAATTTTATGGGACCA 1005

QY 1093 ACTATAGTAACAGGGAAATTTAGTTATAGAGTTGAATATATGACCAT---CAGTCCG 1149
DB 1006 ACTATAGTAACAGGGAAATTTAGTTATAGAGTTGAATATATGACCATCAGCAGTCCG 1065
QY 1150 ATTTTGGATTAACAGCAAAAAGACCTCAAGTTTGGCATTTCACTAACTTAACACCATTTACA 1209
DB 1066 ATTTTGGATTAACAGCAAAAAGACCTCAAGTTTGGCATTTCACTAACTTAACACCATTTACA 1125
QY 1210 ATGTATGATGCTATATTGCGGCTGAAACACAGTGCAGGGACCTGGGCCCAGTCAATATT 1269
DB 1126 ATGTATGATGCTATATTGCGGCTGAAACACAGTGCAGGGACCTGGGCCCAGTCAATATT 1185
QY 1270 TCAGTATTCACTCCACACAGATGTTCCAGGGGAGTGTTCATTTTAACTTGCAGAGGTA 1329
DB 1186 TCAGTATTCACTCCACACAGATGTTCCAGGGGAGTGTTCATTTTAACTTGCAGAGGTA 1245
QY 1330 GAATCCACGCAAGTAGAATTAATCTTGGAAAGAAACCAAGCAACCAAAATGGAATTTAAAC 1389
DB 1246 GAATCCACGCAAGTAGAATTAATCTTGGAAAGAAACCAAGCAACCAAAATGGAATTTAAAC 1305
QY 1390 CAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTGGAAATTAATCTTGCCT 1449
DB 1306 CAATACCGAGTGAAGTGTAGTTCCAGAGACAGGAATTAATTTGGAAATTAATCTTGCCT 1365
QY 1450 ACTGGAATTAATGAGTATATAAATGACCCCATGGCTCCAGAAATTTGGAACATAGTAGAG 1509
DB 1366 ACTGGAATTAATGAG---ATAATGACCCCATGGCTCCAGAAATTTGGAACATAGTAGAG 1422
QY 1510 CCAATGGTAGAATTAATGAGGGTTCCAGAGAGATGCTGTGACCTTCACTCACTTGTCT 1569
DB 1423 CCAATGGTAGAATTAATGAGGGTTCCAGAGAGATGCTGTGACCTTCACTCACTTGTCT 1482
QY 1570 ACATTTATATAACAGCCATCCAGATAAACTTTCCTGCAAGGAATGAGCTGAAGAC 1629
DB 1483 ACATTTATATAACAGCCATCCAGATAAACTTTCCTGCAAGGAATGAGCTGAAGAC 1542
QY 1630 CAGACTTCCACGAGTTGTAATCAAGGAATCAGTATATTAATCACTGACATTCAGCTGAACAG 1689
DB 1543 CAGACTTCCACGAGTTGTAATCAAGGAATCAGTATATTAATCACTGACATTCAGCTGAACAG 1602
QY 1690 CTGCTTATGTTATCAGAGACTTGTACCTTTCACTGAGACATGATGATGTAATCTGCT 1749
DB 1603 CTGACTTATGTTCTTATCAGATTAAGAGATTTTGGGCTGAGACAAATGGGGTTTCTAGA 1662
QY 1750 TTCCACCATCAGGAGGAGGCCAACCAAGTCTCAGTGTAGGACACGTCAGCAAGTG 1809
DB 1663 TATACATCATG-----TCATCTGCAAGCAGGAGCAATTTTGACTTCCCCCAGGC 1710
QY 1810 CCAAGCTCCATTAATAAATTAATAAATAATTAATTAATTTCTTCACTATTTTGTATAT 1869
DB 1711 CTTTGTGAGCCCAAAATTCAGAGTTACATGTTACATTAACAGAAATTTTATACAC 1770
QY 1870 TGGGATCTCCAGAAATPATCCCAATGGAAATAATCACTACATACGATTTATGCAATGGAA 1929
DB 1771 TGGGATCTCCAGATCTCTGATTTTTCATCATTAACCTTATCACTATTTTGGATGTGAA 1830
QY 1930 TTGGATACAAACAGAGCATTCAGATPAACTACCATAGATAACAGCTTCTCATACAGGG 1989
DB 1831 AAACCAATCCAAGAGTATTAATTTTAAGGACATTAACACAGTTTGTCTCTGCTTATAGGG 1890
QY 1990 TTAAGAAATACACAAATACAAATGAGAGTGGCAGGCTCAACCCAGCATGAGAAAGT 2049
DB 1891 TTAAGAAATACACAAATACAAATGAGAGTGGCAGGCTCAACCCAGCTTGGAGAAAGT 1950
QY 2050 TCTTTGTCTGAAAGAAATGACATCTTTGTGAGAACTTTCAGAGATGAACCGGAAATCATCA 2109
DB 1951 TCTTTGTCTGAAAGAAATGACATCTTTGTGAGAACTTTCAGAGATGAACCGGAAATCATCA 2010
QY 2110 CCTCAAGATGTCGAAGTAATGATGTTACCGCAGATGAATTAAGTTGAAGTGTGACCA 2169
DB 2011 CCTCAAGATGTCGAAGTAATGATGTTACCGCAGATGAATTAAGTTGAAGTGTGACCA 2070
QY 2170 CCGGAAAGCCCAATGGGATCATTATTGCTTTATGAGTGTATATAAAAAATATAGATACT 2229

DB 2071 CCGGAAAGCCCAATGGGATCATTATTGCTTATGAAGTGTATATAAAAAATATAGATACT 2130
QY 2230 TTATATATGAAGAACACACATCAACACAGACATATAATTAAGGAACCTTAAGACCTCACACC 2289
DB 2131 TTATATATGAAGAACACATCAACACAGACATATAATTAAGGAACCTTAAGACCTCACACC 2190
QY 2290 CTCTATAACATTTCTGTAAAGTCTTTACACAGATTTGGTCTAGGCAATCAGGTATCTTCT 2349
DB 2191 CTCTATAACATTTCTGTAAAGTCTTTACACAGATTTGGTCTAGGCAATCAGGTATCTTCT 2250
QY 2350 TTACTCTCTGTAAAGACTTTCCGAGACTGTGCTGTATAGTGCACACAGAAATATCATTTAC 2409
DB 2251 TTACTCTCTGTAAAGACTTTCCGAGACTGTGCTGTATAGTGCACACAGAAATATCATTTAC 2310
QY 2410 AAAAATATTTCTTCTGGAGAGATTTGAGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGA 2469
DB 2311 AAAAATATTTCTTCTGGAGAGATTTGAGCTATCATTTCTTCCCCCAAGTAGTCCCAATGGA 2370
QY 2470 ATCATAAAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAAATCGAGAAAGAACTATA 2529
DB 2371 ATCATAAAAAAATATACAAATTTATCTCAAGAGAGTAAATGGAAATCGAGAAAGAACTATA 2430
QY 2530 AATACAAACCTCTTTTAAACCAAAACATTTAAAGTCTGAAGAAATATATACCAATATATCAT 2589
DB 2431 AATACAAACCTCTTTTAAACCAAAACATTTAAAGTCTGAAGAAATATATACCAATATATCAT 2490
QY 2590 GAGGTGTCTGTAGTACACTGAAGGTGAAGGTTCGAGGTGCTCCCAATAGTATACATG 2649
DB 2491 GAGGTGTCTGTAGTACACTGAAGGTGAAGGTTCGAGGTGCTCCCAATAGTATACATG 2550
QY 2650 ACGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAACAGTTGTCTGCT 2709
DB 2551 ACGGAGGAGATGCTCTGATTTCTCCCTCAAGACTTCTCTGTAAACAGTTGTCTGCT 2610
QY 2710 GTCAAGGTGAAGTGTCTGCAACCAACCCCTGGAGCCAAATGGAATATATCTTTATTTAC 2769
DB 2611 GTCAAGGTGAAGTGTCTGCAACCAACCCCTGGAGCCAAATGGAATATATCTTTATTTAC 2670
QY 2770 ACAGTTTATGCTGGAATAGATCANTTAAACATTTAAATGCTACTGAAACATCATTTG 2829
DB 2671 ACAGTTTATGCTGCG---AGATCATCATTTAAACATTTAAATGCTACTGAAACATCATTTG 2727
QY 2830 GAGTTATCAGATTTGGATTTAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACACAGA 2889
DB 2728 GAGTTATCAGATTTGGATTTAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACACAGA 2787
QY 2890 TTTGGTGAATGGAAACAGGAAGCAATATCATTTAGCTTTTCAAAACACAGAGGAGACCA 2949
DB 2788 TTTGGTGAATGGAAACAGGAAGCAATATCATTTAGCTTTTCAAAACACAGAGGG---ACCA 2844
QY 2950 AGCGATCTCCCAAGATGTTTATGCAAACTCAGTTCTTCTCATCAATATTTCTTTTC 3009
DB 2845 AGCGATCTCTCCCAAGATGTTTATGCAAACTCAGTTCTTCTCATCAATATTTCTTTTC 2904
QY 3010 TCGACACCTCTTCAAACTTAAATGGAATTTACAAATTTACTCTGTTTATACAGAAAT 3069
DB 2905 TCGACACCTCTTCAAACTTAAATGGAATTTACAAATTTACTCTGTTTATACAGAAAT 2964
QY 3070 ACTTCAGGTACTTTTATGCAAAATTTTACACTCCATGAACTAAACCAATGCTTTGCAAT 3129
DB 2965 ACTTCAGGTACTTTTATGCAAAATTTTACACTCCATGAACTAAACCAATGCTTTGCAAT 3024
QY 3130 ATGACTGTATCCAAATTTAGATAAATTTAGCAATTTTCACTACTATCATTTTGGTTA 3189
DB 3025 ATGACTGTATCCCAATTTAGATAAATTTAGCAATTTTCACTACTATCATTTTGGTTA 3084
QY 3190 ACAGCAAGTACTTTCAGTTGGAATGGAATAAAGCAGTGACATCAATGGAATATACACA 3249
DB 3085 ACAGCAAGTACTTTCAGTTGGAATGGAATAAAGCAGTGACATCAATGGAATATACACA 3144
QY 3250 GATCAAGACATACCTGAAAGGTTTGTGGAACCTTGAAACCTTGCAATTTGCTTCAACT 3309

Db 3145 GATCAAGAGCTACCTGAAGGTTTGTGGAAACCTGACTTACGAATCCATTTTCGTCAACT 3204
Qy 3310 GCAATAAATGTAAGCTGGGTCCACCGCTCAACCAACGGTCTAGTCTTCTTACTATGTT 3369
Db 3205 GCAATAAATGTAAGCTGGGTCCACCGCTCAACCAACGGTCTAGTCTTCTTACTATGTT 3264
Qy 3370 TCACGTGATCTTACAGAGACTCTCTGCCATGTGAGACCACTCTTGTGTACATATGAGAGA 3429
Db 3265 TCACGTGATCTTACAGAGACTCTCTGCCATGTGAGACCACTCTTGTGTACATATGAGAGA 3324
Qy 3430 AGCATATATTTTGTAAATCTGGAAATACTGTAATATATATTAATAAATTAATCTCCATCA 3489
Db 3325 AGCATATATTTTGTAAATCTGGAAATACTGTAATATATATTAATAAATTAATCTCCATCA 3384
Qy 3490 ACAGAAAAGGGATTTCTGTATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAGAT 3549
Db 3385 ACAGAAAAGGGATTTCTGTATACCTATATCTGCCAGCTATACATCAAGACTGAAGAAGAT 3444
Qy 3550 GTCCAGAAACTTACCAATATCAACACTTTTAAACCTTTCTTACCTCAGTTCTC 3609
Db 3445 ATCCAGAAACTTCAACCAATATCAACACTTTTAAACCTTTCTTACCTCAGTTCTC 3504
Qy 3610 TTATCATGGGATCCCGAGTAAGCCAAATGGTGCATAATAAGTTATGATTTAACTTTA 3669
Db 3505 TTATCATGGGATCCCGAGTAAGCCAAATGGTGCATAATAAGTTATGATTTAACTTTA 3564
Qy 3670 CAAAGACCAATGAAATATTTCTTTCAATTAATCTGTATTAATACATAATATTTGAAGAG 3729
Db 3565 CAAAGACCAATGAAATATTTCTTTCAATTAATCTGTATTAATACATAATATTTGAAGAG 3624
Qy 3730 CTTTCAACATTAATATATAGCTTTTGTGCGGCAAGAACTTAGAAAAGCACTTGGT 3789
Db 3625 CTTTCAACATTAATATATAGCTTTTGTGCGGCAAGAACTTAGAAAAGCACTTGGT 3684
Qy 3790 CTTTCCAGTATCTTTTACACAGATGAGTCACTGCGGTAGCACTTCCACAAAT 3849
Db 3685 CTTTCCAGTATCTTTTACACAGATGAGTCACTGCGGTAGCACTTCCACAAAT 3744
Qy 3850 TTGACTTTAATCAACTGACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTT 3909
Db 3745 TTGACTTTAATCAACTGACTTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCTCTT 3804
Qy 3910 CAGGTGGTATTTGTAAGTATATAGTTTAAATTTCAATGAACATGAACATGACACTATA 3969
Db 3805 CCAGGTGGTATTTGTAAGTATATAGTTTAAATTTCAATGAACATGAACATGACACTATA 3864
Qy 3970 TATTATAGAATATATCAGGATTTAAACTGAAGCAAACTTTGTTGGACTGGAACCACTC 4029
Db 3865 TATTATAGAATATATCAGGATTTAAACTGAAGCAAACTTTGTTGGACTGGAACCACTC 3924
Qy 4030 AGCACTACTCTATCCGTGATCTGCTTTCACAAAGTTGGAATGGCAATCAATTTAGT 4089
Db 3925 AGCACTACTCTATCCGTGATCTGCTTTCACAAAGTTGGAATGGCAATCAATTTAGT 3984
Qy 4090 AATGTAGTAAATTTCAACCAACCAAGATCAGTTCCAGATGTGCGTGCAGATATGCACTGC 4149
Db 3985 AATGTAGTAAATTTCAACCAACCAAGATCAGTTCCAGATGTGCGTGCAGATATGCACTGC 4044
Qy 4150 ATGGCAACTAGCTGGCAGTCACTTTTGTGAAATGGGATCCACCCAAAGGCAATGGA 4209
Db 4045 ATGGCAACTAGCTGGCAGTCACTTTTGTGAAATGGGATCCACCCAAAGGCAATGGA 4104
Qy 4210 ATAATAACGCAATATATGTAACAGTTGAAAGGAATTTCTACAAAGTTTCTCCCAAGAT 4269
Db 4105 ATAATAACGCAATATATGTAACAGTTGAAAGGAATTTCTACAAAGTTTCTCCCAAGAT 4164
Qy 4270 CACATGTACACTTTCAATAAGCTTCTTGCCAAATACCTCATATGTCTTTAAAGTAAGAGCT 4329
Db 4165 CACATGTACACTTTCAATAAGCTTCTTGCCAAATACCTCATATGTCTTTAAAGTAAGAGCT 4224
Qy 4330 TCAACTCAGCTGGTGAAGGTGATGAAGGACATGCCATGTGAGCAGCACTTACCTGAACA 4389
Db 4225 TCAACTCAGCTGGTGAAGGTGATGAAGGACATGCCATGTGAGCAGCACTTACCTGAACA 4284

Qy 4390 GTTCCAGTGTTCACCAAAATATTGCTTTTCTGTATGTTCACTCAACTAGTGCACCAATG 4449
Db 4285 GTTCCAGTGTTCACCAAAATATTGCTTTTCTGTATGTTCACTCAACTAGTGCACCAATG 4344
Qy 4450 ACATGGTAAGACCTGACACTATCTTGGCTACTTTTCAAAATTAACAAATTAACCACTCAA 4509
Db 4345 ACATGGTAAGACCTGACACTATCTTGGCTACTTTTCAAAATTAACAAATTAACCACTCAA 4404
Qy 4510 CTTTGTGTCTCAAAATGCAAAAGATGGGAATCCGAAGATGTTGTTGAATATCAAAAAAT 4569
Db 4405 CTTTGTGTCTCAAAATGCAAAAGATGGGAATCCGAAGATGTTGTTGAATATCAAAAAAT 4464
Qy 4570 CAATACCTCTATGAAGCTCACTTAATCTGAAGAGACAGTATATGGAATTAAGAAAAAT 4629
Db 4465 CAATACCTCTATGAAGCTCACTTAATCTGAAGAGACAGTATATGGAATTAAGAAAAAT 4524
Qy 4630 TGGTATAGATTCCTCAAGTGGCTGCCAGCACCAATGCTGGCTATGGAATGCTTCAAACTGG 4689
Db 4525 TGGTATAGATTCCTCAAGTGGCTGCCAGCACCAATGCTGGCTATGGAATGCTTCAAACTGG 4584
Qy 4690 ATTTCTACAAAACTCTGCTGGCCCTCCAGATGCTCTCTGAAAAATGTTTCATGTAGTA 4749
Db 4585 ATTTCTACAAAACTCTGCTGGCCCTCCAGATGCTCTCTGAAAAATGTTTCATGTAGTA 4644
Qy 4750 GCAACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCTGTCATTTAGTGCACCA 4809
Db 4645 GCAACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCTGTCATTTAGTGCACCA 4704
Qy 4810 ACATGTTATCTGATGATGTCMAATCGGTAGTATATGATGAATTAATATATCTTCTCATC 4869
Db 4705 ACATGTTATCTGATGATGTCMAATCGGTAGTATATGATGAATTAATATATCTTCTCATC 4764
Qy 4870 AAGTCAAAATGAAGAAAAATAAAACCATAGAAATTAAGATTTTAAAGATTTTCACAAGGTAT 4929
Db 4765 AAGTCAAAATGAAGAAAAATAAAACCATAGAAATTAAGATTTTAAAGATTTTCACAAGGTAT 4824
Qy 4930 TCTGTAGTGAATCAGTGCATTTTCTGGGAACATTTAGTGTGTCATATGTAGAAGGAGTCA 4989
Db 4825 TCTGTAGTGAATCAGTGCATTTTCTGGGAACATTTAGTGTGTCATATGTAGAAGGAGTCA 4884
Qy 4990 AGTGTGAAATGATGTTTACTACTTTAGATCAGCCCCCAAGGACCCACCTTAACAACATG 5049
Db 4885 AGTGTGAAATGATGTTTACTACTTTAGATCAGCCCCCAAGGACCCACCTTAACAACATG 4944
Qy 5050 ACATTTCAAGAGATACAGAGTGAAGTTACAAAATTTCAATTAACGTTCTCTCTCTCT 5109
Db 4945 ACATTTCAAGAGATACAGAGTGAAGTTACAAAATTTCAATTAACGTTCTCTCTCTCT 5004
Qy 5110 CAACCTAATGGAANAATATCAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCCTACT 5169
Db 5005 CAACCTAATGGAANAATATCAAGTATATCAAGCTCTGGTTTACCGAGAAGATGATCCTACT 5064
Qy 5170 GCTGTCCAGATTCACAACTCAGATTTATACAGAAAACCAACACATTTGTCATTTGCAATG 5229
Db 5065 GCTGTCCAGATTCACAACTCAGATTTATACAGAAAACCAACACATTTGTCATTTGCAATG 5124
Qy 5230 CTAGAAGCACTAAAGGTTGGACATACATCAATATCAGTGTGTTACGCACTCAATAGTGT 5289
Db 5125 CTAGAAGCACTAAAGGTTGGACATACATCAATATCAGTGTGTTACGCACTCAATAGTGT 5184
Qy 5290 GGTGAGGTCTCAAAAGGTTCCGATGAGAAATAACCATGGATATCAAAAGCTCCAGCAGCA 5349
Db 5185 GGTGAGGTCTCAAAAGGTTCCGATGAGAAATAACCATGGATATCAAAAGCTCCAGCAGCA 5244
Qy 5350 AAAACCAAAACCAACCCCTATTTATGATGCCACAGAAAATCTGCTGTGACATTTCAACA 5409
Db 5245 AAAACCAAAACCAACCCCTATTTATGATGCCACAGAAAATCTGCTGTGACATTTCAACA 5304
Qy 5410 ATTACAAATCAGATGCCCAATATGTTACTACATGATGATCATGGACCAATTAATAATGTA 5469
Db 5305 ATTACAAATCAGATGCCCAATATGTTACTACATGATGATCATGGACCAATTAATAATGTA 5364

QY 5470 CAAGTGTTCGACAGAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGAT 5529
DB 5365 CAAGTGTTCGACAGAACAGGAGCTCAGCATGATGGAATGTAAACAAAGTGGTATGAT 5424
QY 5530 GCATATTTTAAATAAGCAAGGCCATATTTTAAATGAAGCTTTCTTAAACCTCCATGT 5589
DB 5425 GCATATTTTAAATAAGCAAGGCCATATTTTAAATGAAGCTTTCTTAAACCTCCATGT 5484
QY 5590 ACAGAGGAAGACAAAGCTTTAGTGGCAATGAGGAATCTACATCATAGTGGCTGATTAAT 5649
DB 5485 ACAGAGGAAGACAAAGCTTTAGTGGCAATGAGGAATCTACATCATAGTGGCTGATTAAT 5544
QY 5650 GCATGCATGATTCCTGGCAATGAGCAAAATTTTGCATGGAACCACTGAAACCAAAAG 5709
DB 5545 GCATGCATGATTCCTGGCAATGAGCAAAATTTTGCATGGAACCACTGAAACCAAAAG 5604
QY 5710 CAATACTTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACCTGATTAAT 5769
DB 5605 CAATACTTATTTAAATTTAGAGCTACAAATATTTATGGGCAATTTTACCTGATTAAT 5664
QY 5770 TCTGACCTGTATAGACTTTAGGGGAAGGACTTTTGCAGAAAGAACCGTAGAGATCATTTCT 5829
DB 5665 TCTGACCTGTATAGACTTTAGGGGAAGGACTTTTGCAGAAAGAACCGTAGAGATCATTTCT 5724
QY 5830 TCCGTCACCTTTGTGTATCTTTCAATAATTTCTCTTGGAAACAGCTATTTTGGCAATTTGCA 5889
DB 5725 TCCGTCACCTTTGTGTATCTTTCAATAATTTCTCTTGGAAACAGCTATTTTGGCAATTTGCA 5784
QY 5890 AGAATTCGACAGAGCAGAAAGAGTGGCAATCTCTCTCAGGATGCGAGAAATTAAT 5949
DB 5785 AGAATTCGACAGAGCAGAAAGAGTGGCAATCTCTCTCAGGATGCGAGAAATTAAT 5844
QY 5950 GACACTAAATTTAGAGCTGATCAGCTCATCAGTGGCAGACCTGGAACGAG 6009
DB 5845 GACACTAAATTTAGAGCTGATCAGCTCATCAGTGGCAGACCTGGAACGAG 5904
QY 6010 AGATTAAACGGCGCAATAAGCAAGAAATCTTCTCGCAACATGTTGAAGAGCTTTGCACA 6069
DB 5905 AGATTAAACGGCGCAATAAGCAAGAAATCTTCTCGCAACATGTTGAAGAGCTTTGCACA 5964
QY 6070 AACCAACACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTCCCAAAATTTCTTCAGGATCTT 6129
DB 5965 AACCAACACCTAAAGTTTCAAGAGAAATTTTCGGAAATTTCCCAAAATTTCTTCAGGATCTT 6024
QY 6130 TCTTCAACTGATGCTGATCTGCTGCTGGAATAGAGCAAAACCGTTTCCCAAAACATAAA 6189
DB 6025 TCTTCAACTGATGCTGATCTGCTGCTGGAATAGAGCAAAACCGTTTCCCAAAACATAAA 6084
QY 6190 CCATATAATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCACAGTTCCGAT 6249
DB 6085 CCAT---ATAATAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCACAGTTCCGAT 6141
QY 6250 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCCAAATGAATTTATTTGCTACTCAA 6309
DB 6142 TATATTAATGCCAGCTATATTTCTGGTATTTATGTCCAAATGAATTTATTTGCTACTCAA 6201
QY 6310 GGTCCACTACCAAGAACAGTTGGAGATTTTGGAGAAATGGTGGGAACACAGGGCAAAA 6369
DB 6202 GGTCCACTACCAAGAACAGTTGGAGATTTTGGAGAAATGGTGGGAACACAGGGCAAAA 6261
QY 6370 ACATTAGTAATGCTAAACAGTGTTTTGAAGAGGAGGATCAGATGCCATCAGTATTGG 6429
DB 6262 ACATTAGTAATGCTAAACAGTGTTTTGAAGAGGAGGATCAGATGCCATCAGTATTGG 6321
QY 6430 CCAGAGGACAAACAGCCAGTTACTCTCTTTGGAGATATAGTGATTTACAAAGCTAATGGAG 6489
DB 6322 CCAGAGGACAAACAGCCAGTTACTCTCTTTGGAGATATAGTGATTTACAAAGCTAATGGAG 6381
QY 6490 GATGTTCAAAATAGATTGCACTATCAGGATCTGAAATTTGAAGCATGGGGATTTGCATG 6549
DB 6382 GATGTTCAAAATAGATTGCACTATCAGGATCTGAAATTTGAAGCATGGGGATTTGCATG 6441
QY 6550 ACTGTTTCAGAGTGTAACTTTTACTGCCTGGCCAGAGCATGGGTTCTCTGAGAACAGCGCC 6609

DB 6442 ACTGTTTCAGAGTGTAACTTTTACTGCTGGCCAGAGCATGGGTTCTCTGAGAACAGCGCC 6501
QY 6610 CCTCTAATTCACCTTTGTGAAGTTGCTTCGACAGCAGGCGACATGACACACACCTATG 6669
DB 6502 CCTCTAATTCACCTTTGTGAAGTTGCTTCGACAGCAGGCGACATGACACACACCTATG 6561
QY 6670 ATTGTTCACTGCACTGCTGGAGTTGGAAGAACTCGAGTTTTTATTTGCTCTCGACCATTTA 6729
DB 6562 ATTGTTCACTGCACTGCTGGAGTTGGAAGAACTCGAGTTTTTATTTGCTCTCGACCATTTA 6621
QY 6730 ACACAACATATAAATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACCTGAGA 6789
DB 6622 ACACAACATATAAATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACCTGAGA 6681
QY 6790 AGTGAAGAATGTGCATGGTGCAGAACTGCGCAGTATATCTTTTACACCACTGCAAT 6849
DB 6682 AGTGAAGAATGTGCATGGTGCAGAACTGCGCAGTATATCTTTTACACCACTGCAAT 6741
QY 6850 CTGGATCTCTTATCAAAATTAAGGGAAGTAATCAGGCCCATCTGTTTGTAACTATTTCAGCA 6909
DB 6742 CTGGATCTCTTATCAAAATTAAGGGAAGTAATCAGGCCCATCTGTTTGTAACTATTTCAGCA 6801
QY 6910 CTTCAGAAGATGGAATCTTTTGGACGCCATGGAA--GGTGAATGTGAGCTTGAATGGAA 6966
DB 6802 CTTCAGAAGATGGAATCTTTTGGACGCCATGGAAAGGTGGTGAATGTGAGCTTGAATGGAA 6861
QY 6967 GAAACCATATGTAATAATTTACAGACCAAGATACAAATGGAAGAGATTTTAAATCCCA 7026
DB 6862 GAAACCATATGTAATAATTTACAGACCAAGATACAAATGGAAGAGATTTTAAATCCCA 6921
QY 7027 GGGGCCAAAGTTACCCCTCATCTTCGGAATGAAATGTCAACCTTAAAGAAATATCT 7086
DB 6922 GGGGCCAAAGTTACCCCTCATCTTCGGAATGAAATGTCAACCTTAAAGAAATATCT 6981
QY 7087 ATGCTTCTCTCAC 7099
DB 6982 ATGCTTCTCTCAC 6994

RESULT 8

US-10-314-232-21
; Sequence 21, Application US/10314232
; Publication No. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 21
; LENGTH: 3973
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)..(3973)
; OTHER INFORMATION: n = unknown nucleotide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2694)

US-10-314-232-21

Query Match 36.8%; Score 2618.6; DB 15; Length 3973;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2716; Conservative 0; Mismatches 4; Indels 75; Gaps 2;

QY 4386 AACAGTCCAGTGTCCACAAATATGCTTTTCTGATGTTCAAGTCAACTAGTGCAC 4445
DB 33 AAAAGTTCACAGTGTCCACAAATATGCTTTTCTGATGTTCAAGTCAACTAGTGCAC 92
QY 4446 ATTGACATGATGAAGACCTGACACTATCTCTGGCTACTTTCAAAATTTACAAATTTACCAC 4505
DB 93 ATTGACATGATGAAGACCTGACACTATCTCTGGCTACTTTCAAAATTTACAAATTTACCAC 152
QY 4506 TCAACTCTGCTCAAAAATGCAAGAATGGAAATCGAAGATGTGTTGAATATCAAAA 4565
DB 153 TCAACTCTGCTCAAAAATGCAAGAATGGAAATCGAAGATGTGTTGAATATCAAAA 212
QY 4566 AATTCATACCTCTATGAAGCTCACTTAATCTGAAGACAGATATATGGATTAAGAAATTT 4625
DB 213 AATTCATACCTCTATGAAGCTCACTTAATCTGAAGACAGATATATGGATTAAGAAATTT 272
QY 4626 TAGATGTTATAGATTCGAAGTGGCTGCGACCAATGCTGGCTATGGCAATGTTCAAA 4685
DB 273 TAGATGTTATAGATTCGAAGTGGCTGCGACCAATGCTGGCTATGGCAATGTTCAAA 332
QY 4686 CTGGATTTCTCAAAAATCTCTGCTGGCCCTCCAGATGCTCTCTGAAAATGTTCAATGT 4745
DB 333 CTGGATTTCTCAAAAATCTCTGCTGGCCCTCCAGATGCTCTCTGAAAATGTTCAATGT 392
QY 4746 AGTAGCAATCACTTTTAGCATCAGCATAGCTGAGTGAACCTGCTGCTCAATTTACTGG 4805
DB 393 AGTAGCAATCACTTTTAGCATCAGCATAGCTGAGTGAACCTGCTGCTCAATTTACTGG 452
QY 4806 ACCACATGTTATCTGATGATGTCAAATCGGTAGTAAATGATGATTAATATATCTTT 4865
DB 453 ACCACATGTTATCTGATGATGTCAAATCGGTAGTAAATGATGATTAATATATCTTT 512
QY 4866 CATCAAGTCAAAATGAAGAAATTAACCAATAGAAATTAAGAAATTTAGAAATTTCAACAG 4925
DB 513 CATCAAGTCAAAATGAAGAAATTAACCAATAGAAATTTAGAAATTTCAACAG 572
QY 4926 GTATTTCTGATGATCACTGCAATTTAATGCGAATTTAGTGTGCTATGTTAGAGGGAA 4985
DB 573 GTATTTCTGATGATCACTGCAATTTAATGCGAATTTAGTGTGCTATGTTAGAGGGAA 632
QY 4986 GTCAAGTGTGAAATGATGTTACTACTTTAGAAATCAGCCCAAGAGCCCACTTAACAA 5045
DB 633 GTCAAGTGTGAAATGATGTTACTACTTTAGAAATCAGCCCAAGAGCCCACTTAACAA 692
QY 5046 CATGACATTTCAAGAGATACCAAGTGAAGTAAATTTCAAAATTTCAAAATTTCTCTCC 5105
DB 693 CATGACATTTCAAGAGATACCAAGTGAAGTAAATTTCAAAATTTCAAAATTTCTCTCC 752
QY 5106 TTCTCAACCTAAATGGAATATATCAAGTATATCAAGTCTGTTTACCGAAGATGATCC 5165
DB 753 TTCTCAACCTAAATGGAATATATCAAGTATATCAAGTCTGTTTACCGAAGATGATCC 812
QY 5166 TACTGCTGTCCAGATTTCAACCTCAATTTATACAGAAACCAACATTTGCTCAATGTC 5225
DB 813 TACTGCTGTCCAGATTTCAACCTCAATTTATACAGAAACCAACATTTGCTCAATGTC 872
QY 5226 AATGCTAGAGGACTAAAGGTGGACATACATACATATCACTGTTTACGAGTCAATAG 5285
DB 873 AATGCTAGAGGACTAAAGGTGGACATACATACATATCACTGTTTACGAGTCAATAG 932
QY 5286 TGCTGTGTCCAGGTCCAAAGTTCGATGAGAAATACCAATGATATCAAGCTCCAGCAG 5345
DB 933 TGCTGTGTCCAGGTCCAAAGTTCGATGAGAAATACCAATGATATCAAGCTCCAGCAG 992
QY 5346 ACCAAAAACCAACCAACCCCTATTTATGATGCCACAGAAATCTGCTTGTGACTTCAAC 5405
DB 993 ACCAAAAACCAACCAACCCCTATTTATGATGCCACAGAAATCTGCTTGTGACTTCAAC 1052

QY 5406 AACAAATTCAATCAGAATGCCAAATATGTTACTATAGTATGATCATGGAACCAATAAAAA 5465
DB 1053 AACAAATTCAATCAGAATGCCAAATATGTTACTATAGTATGATCATGGAACCAATAAAAA 1112
QY 5466 TGTCAAGTCTGCTGCGACAGAAACAGGAGCTCAGCATGATGGAATGTAACAAAGGTGA 5525
DB 1113 TGTCAAGTCTGCTGCGACAGAAACAGGAGCTCAGCATGATGGAATGTAACAAAGGTGA 1172
QY 5526 TGATGCATATTTTAATAAGCAAGGCCATATTTTCAAAATGAAGCTTTCTTAACCCCTCC 5585
DB 1173 TGATGCATATTTTAATAAGCAAGGCCATATTTTCAAAATGAAGCTTTCTTAACCCCTCC 1232
QY 5586 ATGTAACAGAAAGAAAGACAAAGTTTGTGGCAATGAAGAAATCTACATCATATAGTGTCTGA 5645
DB 1233 ATGTAACAGAAAGAAAGACAAAGTTTGTGGCAATGAAGAAATCTACATCATATAGTGTCTGA 1292
QY 5646 TAATGCATGCTGATTTCTGCGCAATGAAGCAAAATTTGCAATGACCACTGCAACCAAA 5705
DB 1293 TAATGCATGCTGATTTCTGCGCAATGAAGCAAAATTTGCAATGACCACTGCAACCAAA 1352
QY 5706 AAAGCAATATCTTATTTAAATTTAGAGCTACAAATATATGGAACAAATTTTACTGACTCTGA 5765
DB 1353 AAAGCAATATCTTATTTAAATTTAGAGCTACAAATATATGGAACAAATTTTACTGACTCTGA 1412
QY 5766 TTATTTCTGACCTGTTAAAGCTTTTAGGGGAAGGACTTTTCAAGAAAGAAACCGTAGAGATCAT 5825
DB 1413 TTATTTCTGACCTGTTAAAGCTTTTAGGGGAAGGACTTTTCAAGAAAGAAACCGTAGAGATCAT 1472
QY 5826 TCTTTCCGCTCACTTTGTGTATCTTTCAATTAATTTCTCTGGAACAGCTATTTTGTGCAAT 5885
DB 1473 TCTTTCCGCTCACTTTGTGTATCTTTCAATTAATTTCTCTGGAACAGCTATTTTGTGCAAT 1532
QY 5886 TGCAGAAATTTGCAGAGAGCAAGAAAGGTGCGACATCTCTCTCAGAGTGCAGAAAT 5945
DB 1533 TGCAGAAATTTGCAGAGAGCAAGAAAGGTGCGACATCTCTCTCAGAGTGCAGAAAT 1592
QY 5946 TATTGACACTAAATTTGAAGCTGGATCAGCTCATCATCAGTGGCAGACCTGGAACCTGAAGGA 6005
DB 1593 TATTGACACTAAATTTGAAGCTGGATCAGCTCATCATCAGTGGCAGACCTGGAACCTGAAGGA 1652
QY 6006 CGAGAGATTAACGG----- 6020
DB 1653 CGAGAGATTAACGGCGATATCTCTTCAATTTTCTTTAGACGCAAGGAGATTTTGTCTATCCA 1712
QY 6021 -----GCCAATAAGCAAGAAATCTCTCTCCTCCTCAACATCT 6053
DB 1713 GTTACTTATAGTAAATTCATCAAGCCAAATAGCAAGAAATCTCTCTCCTCAACATCT 1772
QY 6054 TGAAGAGCTTTTGCAAAACCAACCTAAAGTTTCAAGAAAGAAATTTTTCGAAATTTACCAAA 6113
DB 1773 TGAAGAGCTTTTGCAAAACCAACCTAAAGTTTCAAGAAAGAAATTTTTCGAAATTTACCAAA 1832
QY 6114 ATTTCTTCAAGATCTTTCTTCAATGTATGTCTGCTGGAATAGAGCAAAACCG 6173
DB 1833 ATTTCTTCAAGATCTTTCTTCAATGTATGTCTGCTGGAATAGAGCAAAACCG 1892
QY 6174 TTTCCCAACCAATAAACCATATATAATTAACAGAGTAAAGCTGATAGCTGACGCTAG 6233
DB 1893 CTTCCTCAACCAATAAACCATATATAATTAACAGAGTAAAGCTGATAGCTGACGCTAG 1949
QY 6234 TGTTCAGGTTCCGATTTATATTAATGCGCACTATATTTCTGTTATTTATGTTCCAAATCA 6293
DB 1950 TGTTCAGGTTCCGATTTATATTAATGCGCACTATATTTCTGTTATTTATGTTCCAAATCA 2009
QY 6294 ATTTATTTCTACTCAAGGTTCACCTACAGGAACAGTTGAGAGATTTTTCGAGAAATGCTGTG 6353
DB 2010 ATTTATTTCTACTCAAGGTTCACCTACAGGAACAGTTGAGAGATTTTTCGAGAAATGCTGTG 2069
QY 6354 GGAACACAGGCAAAACATTTAGTAACTGCAAGTGTGTTTGAAGAGCGATCAG 6413
DB 2070 GGAACACAGGCAAAACATTTAGTAACTGCAAGTGTGTTTGAAGAGCGATCAG 2129

```
QY 6414 ATGCCATCATGTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTAT 6473
Db 2130 ATGCCATCATGTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTAT 2189
QY 6474 TACAAAGCTAATGGAGGATGTTCAAATAGATTGGACTATCAGGGATCTGAAAAATTGAAAG 6533
Db 2190 TACAAAGCTAATGGAGGATGTTCAAATAGATTGGACTATCAGGGATCTGAAAAATTGAAAG 2249
QY 6534 GCATGGGATTCATGACCTGTTGCGACAGTGAACCTTTACTGCTGGCCAGACATGGGT 6593
Db 2250 GCATGGGATTCATGACCTGTTGCGACAGTGAACCTTTACTGCTGGCCAGACATGGGT 2309
QY 6594 TCTCAGAACACGGCCCTCTAATTCACCTTTGGAAGTTGGTTTCAGAGCAAGCAGGCGACA 6653
Db 2310 TCTCAGAACACGGCCCTCTAATTCACCTTTGGAAGTTGGTTTCAGAGCAAGCAGGCGACA 2369
QY 6654 TGACACCACTATGATTGTTTCACTGCACTGCTGGAGTTGGAAGAACTGGAGTTTAT 6713
Db 2370 TGACACCACTATGATTGTTTCACTGCACTGCTGGAGTTGGAAGAACTGGAGTTTAT 2429
QY 6714 TGCTCTGGACCAATTTAAACACACATATAAATGACCATGATTTTGGATATATATGGACT 6773
Db 2430 TGCTCTGGACCAATTTAAACACACATATAAATGACCATGATTTTGGATATATATGGACT 2489
QY 6774 AGTAGCTGAATGAGAGTGAAGAAATGTGCATGTGCGAATCTGGCACTATATATCTT 6833
Db 2490 AGTAGCTGAATGAGAGTGAAGAAATGTGCATGTGCGAATCTGGCACTATATATCTT 2549
QY 6834 TTTCACCACTGCTGCTGATCTCTTATCAATTAAGGAACTAATCAGCCCATCTGTTT 6893
Db 2550 TTTCACCACTGCTGCTGATCTCTTATCAATTAAGGAACTAATCAGCCCATCTGTTT 2609
QY 6894 TGTTAACTATTCAGCACTTCAGAACATGCACTCTTTGGACGCAATGGAGTGAATGGA 6953
Db 2610 TGTTAACTATTCAGCACTTCAGAACATGCACTCTTTGGACGCAATGGAGTGAATGGA 2669
QY 6954 GCTTGAATGGGAGAAACACATATGTAATATTTACAGCAAGGATACAAATGGGAAGAGA 7013
Db 2670 GCTTGAATGGGAGAAACACATATGTAATATTTACAGCAAGGATACAAATGGGAAGAGA 2729
QY 7014 TTTTAAATCCAGGGGCAAAAGTTACCCCTCATCTTCCGAATGGAATGCAACCT 7073
Db 2730 TTTTAAATCCAGGGGCAAAAGTTACCCCTCATCTTCCGAATGGAATGCAACCT 2789
QY 7074 TAAAGAAATATATGCTTCTCTCACTGCGCTTT 7108
Db 2790 TAAAGAAATATATGCTTCTCTCACTGCGCTTT 2824
```

RESULT 9

US-10-087-684-7

; Sequence 7, Application US/10087684

; Publication No. US20040029116A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: MacDougall, John R.

; APPLICANT: Millet, Isabelle

; APPLICANT: Ellerman, Karen

; APPLICANT: Stone, David J.

; APPLICANT: Grosse, William M.

; APPLICANT: Lepley, Denise M.

; APPLICANT: Rieger, Daniel K.

; APPLICANT: Burgess, Catherine E.

; APPLICANT: Casman, Stacie, J.

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Boldog, Ferenc L.

; APPLICANT: Li, Li

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Mishra, Vishnu

; APPLICANT: Shenoy, Suresh G.

; APPLICANT: Rastelli, Luca

; APPLICANT: Tchernov, Velizar T.

; APPLICANT: Vernet, Corine A.M.

```
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Esha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 7
; LENGTH: 2565
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2565)
US-10-087-684-7
```

Query Match 35.9%; Score 2551; DB 17; Length 2565;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2554; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 3262 CCTGAAGGGTTGTTGGAAACCTGACTTACGAATCCATTTCTGTCACATGCAATAAATGTA 3321
Db 7 CCTGAAGGGTTGTTGGAAACCTGACTTACGAATCCATTTCTGTCACATGCAATAAATGTA 66
QY 3322 AGCTGGGTCCCAACCGGCTCAACCAACCGGTCTAGTCTTCTACTATGTTTCACTGATCTTA 3381
Db 67 AGCTGGGTCCCAACCGGCTCAACCAACCGGTCTAGTCTTCTACTATGTTTCACTGATCTTA 126
QY 3382 CAGCAGACTCTCGCCATGTGAGCACCCTCTGTTACATATGAGAGAGCATATATTTT 3441
Db 127 CAGCAGACTCTCGCCATGTGAGCACCCTCTGTTACATATGAGAGAGCATATATTTT 186
QY 3442 GATATCTGGAAATAATACACTGATTATATATATAAATAATTACTCCATCAACAGAAAGGA 3501
Db 187 GATATCTGGAAATAATACACTGATTATATATAAATAATTACTCCATCAACAGAAAGGA 246
QY 3502 TTCTCTGATACCTATACCTGCCAGCTATACATCAAGACTGAAGAAGATGTCCCAAAACT 3561
Db 247 TTCTCTGATACCTATACCTGCCAGCTATACATCAAGACTGAAGAAGATGTCCCAAAACT 306
QY 3562 TCACCAATAATCAACACTTTTAAAAACCTTCTCTACTAGTCTCTTATCATGGGAT 3621
Db 307 TCACCAATAATCAACACTTTTAAAAACCTTCTCTACTAGTCTCTTATCATGGGAT 366
QY 3622 CCCCAGTAAAGCCAAATGGTGCAATAAATAGTTATGATTTAACTTTAAAGGACCAAT 3681
Db 367 CCCCAGTAAAGCCAAATGGTGCAATAAATAGTTATGATTTAACTTTAAAGGACCAAT 426
QY 3682 GAAATATTTCTTTCAATTAATTTGATTAATAATAATAATAATAATAATAATAATAATAATA 3741
Db 427 GAAATATTTCTTTCAATTAATTTGATTAATAATAATAATAATAATAATAATAATAATAATA 486
QY 3742 ACATATATAGCTTTTCTGCGCCAGCAAGCACTAGAAAGCACTTGGTCTTCCAGTAT 3801
Db 487 ACATATATAGCTTTTCTGCGCCAGCAAGCACTAGAAAGCACTTGGTCTTCCAGTAT 546
QY 3802 CTTTCTTTTACAGAGATGAGTGCCTGCGCTTAGCACCTCCACAAAATTTGACTTTAATC 3861
```


Db 1447 ACTCTGCTGGCCCTCCAGATGCTCTCTGAAATGTTTCATGTAGTACCAATCACT 1506
Qy 4762 TTTAGCATCAGCATAGCTGGAGTGAACTCTGTGTCATTAATCGAGCAACATGTTATCTG 4821
Db 1507 TTTAGCATCAGCATAGCTGGAGTGAACTCTGTGTCATTAATCGAGCAACATGTTATCTG 1566
Qy 4822 ATTGATGTCAAATCGGTAGATGATGATTAATTAATATATCTTTCATCAAGTCAATGAA 4881
Db 1567 ATTGATGTCAAATCGGTAGATGATGATTAATTAATATATCTTTCATCAAGTCAATGAA 1626
Qy 4882 GAAATATAAACCATAGAAATTTAAAGATTTTAGAAATATTCACAAGGTATTTCTGTAGTGATC 4941
Db 1627 GAAATATAAACCATAGAAATTTAAAGATTTTAGAAATATTCACAAGGTATTTCTGTAGTGATC 1686
Qy 4942 ACTGCATTTACTGGGAACATTTAGTGTGTCATATGTAGAGGGAAGTCAAGTGTGGAATG 5001
Db 1687 ACTGCATTTACTGGGAACATTTAGTGTGTCATATGTAGAGGGAAGTCAAGTGTGGAATG 1746
Qy 5002 ATTGTTACTACTTTAGAAATCAGCCCAAGAGCCACCTTAACACATGATCATTTCCAGAG 5061
Db 1747 ATTGTTACTACTTTAGAAATCAGCCCAAGAGCCACCTTAACACATGATCATTTCCAGAG 1806
Qy 5062 ATACCATGATCAAGTTACAAAATTTCAATTAAGCTTCTCTCTCTCTCAACCTAATGGA 5121
Db 1807 ATACCATGATCAAGTTACAAAATTTCAATTAAGCTTCTCTCTCTCAACCTAATGGA 1866
Qy 5122 AATATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTCTACTCTGTCCAGATT 5181
Db 1867 AATATCCAAGTATATCAAGCTCTGTTTACCGAGAAGATGATCTCTACTCTGTCCAGATT 1926
Qy 5182 CACACCTCAGTATATACAGAAACCAACACATTCGTTCATTCGTAATGCTAGAGACTA 5241
Db 1927 CACACCTCAGTATATACAGAAACCAACACATTCGTTCATTCGTAATGCTAGAGACTA 1986
Qy 5242 AAGGTGGACATACATACATATACGTGTTTACGAGTCAATAGTGTCTGTGCTGCTCCA 5301
Db 1987 AAGGTGGACATACATACATATACGTGTTTACGAGTCAATAGTGTCTGTGCTGCTCCA 2046
Qy 5302 AAGGTTCGATGAGATATACCATGATATCAAGCTTCCAGACGACCAAAACCAACCA 5361
Db 2047 AAGGTTCGATGAGATATACCATGATATCAAGCTTCCAGACGACCAAAACCAACCA 2106
Qy 5362 ACCCTATTTATGATGCCACAGAAACCTGTTTGACATTCACCAACAAATTAATCAGA 5421
Db 2107 ACCCTATTTATGATGCCACAGAAACCTGTTTGACATTCACCAACAAATTAATCAGA 2166
Qy 5422 ATGCCAATATGTTACTACAGTATGATCATGACCAATTAATAAATCTACAAGTGTGCG 5481
Db 2167 ATGCCAATATGTTACTACAGTATGATCATGACCAATTAATAAATCTACAAGTGTGCG 2226
Qy 5482 ACAGAAACAGGAGCTCAGCATGATGAAATGTAAACAAAGTGTATGATGATATTTAAT 5541
Db 2227 ACAGAAACAGGAGCTCAGCATGATGAAATGTAAACAAAGTGTATGATGATATTTAAT 2286
Qy 5542 AAGCAAGGCCATATTTTACAAATGAAGCTTTCTTAACCTCCATGTACAGAGGAAAG 5601
Db 2287 AAGCAAGGCCATATTTTACAAATGAAGCTTTCTTAACCTCCATGTACAGAGGAAAG 2346
Qy 5602 ACAGATTTAGTGGCAATCAAGAAATCTACATCATAGTGTGATGATGATGATGATGAT 5661
Db 2347 ACAGATTTAGTGGCAATCAAGAAATCTACATCATAGTGTGATGATGATGATGATGAT 2406
Qy 5662 CTGGCAATGAAGCAAAATTTGCAATGACCACTGAAACCAAAAGCAATACCTATTTT 5721
Db 2407 CTGGCAATGAAGCAAAATTTGCAATGACCACTGAAACCAAAAGCAATACCTATTTT 2466
Qy 5722 AAATTTAGAGCTACAAATATTTAGGACAAATTTACTGATCTGATATTTCTGACCTGTT 5781
Db 2467 AAATTTAGAGCTACAAATATTTAGGACAAATTTACTGATCTGATATTTCTGACCTGTT 2526
Qy 5782 AAGACTTTAGGGGAGGACTTTTCAGAAAGAACCGTAGAG 5820
Db 2527 AAGACTTTAGGGGAGGACTTTTCAGAAAGAACCGTAGAG 2565

RESULT 11

US-10-314-232-14
; Sequence 14, Application US/10314232
; Publication No. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)..(2692)
; OTHER INFORMATION: n = unknown nucleotide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (92)..139, 259..1414)
US-10-314-232-14

Query Match 15.7%; Score 1118.8; DB 15; Length 2692;

Best Local Similarity 93.7%; Pred. No. 7.6e-256;

Matches 1218; Conservative 0; Mismatches 7; Indels 75; Gaps 2;

Qy 5881 GCATTTGCAAGAAATTCGACAGAGCAGAAAGAGGTGGCACATATCTCTCTCAGGATGCA 5940
Db 247 GCTGTTTGTAGATTCGACAGAGCAGAAAGAGGTGGCACATATCTCTCTCAGGATGCA 306
Qy 5941 GAAATATTGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTG 6000
Db 307 GAAATATTGACACTAAATTTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTG 366
Qy 6001 AAGGACGAGAGATTAACCGG----- 6020
Db 367 AAGGACGAGAGATTAACCGGATCTCTTCATTTTCTTTAGACGCAAGGAGATTTTGTGTC 426
Qy 6021 -----GCCAATAAGCAAGAAATCTCTTCCTGCAA 6048
Db 427 ATCCAGTTACTTAGTTATAGAAATCCATCAAGCCCAATAGCAAGAAATCTCTTCCTGCAA 486
Qy 6049 CATGTTGAAGAGCTTTGCAACAAACAACTTAAAGTTTCAAGAAAGAAATTTTCGGAATTA 6108
Db 487 CATGTTGAAGAGCTTTGCAACAAACAACTTAAAGTTTCAAGAAAGAAATTTTCGGAATTA 546
Qy 6109 CCAAAATTTCTTCAGGATCTTTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAA 6168
Db 547 CCAAAATTTCTTCAGGATCTTTTCAACTGATGCTGATCTGCTTGGAAATAGAGCAAAA 606
Qy 6169 AACGTTTCCCAACATAAACCATATATTAATTAACAGATGAAGCTGATAGCTGAC 6228
Db 607 AACGTTTCCCAACATAAACCCAT---ATAATATAACAGATGAAGCTGATAGCTGAC 663
Qy 6229 GCTAGTGTTCAGGTTTCGGATTTATTAATGCGAGCTATATTTCTGTTTATTTATGTCGA 6288
Db 664 GCTAGTGTTCAGGTTTCGGATTTATTAATGCCAGCTATATTTCTGGTATTTATGTCGA 723
Qy 6289 AATGAATTTATGCTACTCAAGGTCCTACCTACAGAACAGTGTGGAGATTTTGGAGAAATG 6348

Db 724 AATGAATTTATTGCTACTCAAGTCCACTACCAGGAACAGTTGGAGATTTTGGAGATG 783
Qy 6349 GTGTGGGAACCAAGGCAAAACATTTAGTAATGCTTAACACAGTGTTTGAAAAGGACGG 6408
Db 784 GTGTGGGAACCAAGGCAAAACATTTAGTAATGCTTAACACAGTGTTTGAAAAGGACGG 843
Qy 6409 ATCAGATGCCATCAGTATTGGCCAGAGGCAACAAGCCAGTTACTGTCTTTGGAGATATA 6468
Db 844 ATCAGATGCCATCAGTATTGGCCAGAGGCAACAAGCCAGTTACTGTCTTTGGAGATATA 903
Qy 6469 GTGATTACAAAGCTAATGAGAGATGTTCAAATAGATGGAATCAGAGGATCTGAAAAT 6528
Db 904 GTGATTACAAAGCTAATGAGAGATGTTCAAATAGATGGAATCAGAGGATCTGAAAAT 963
Qy 6529 GAAAGCATGGGATGTCATGCTGTCACAGTGTAACTTTACTGCGCCGACAGCAT 6588
Db 964 GAAAGCATGGGATGTCATGCTGTCACAGTGTAACTTTACTGCGCCGACAGCAT 1023
Qy 6589 GGGGTTTCTGAGAACAGCGCCCTCTAAATTTCACTTTGTAAGTTGGTTGCGACAGCAGG 6648
Db 1024 GGGGTTTCTGAGAACAGCGCCCTCTAAATTTCACTTTGTAAGTTGGTTGCGACAGCAGG 1083
Qy 6649 GCACATGACACACACTATGATTTGTTCACTGAGTGTGGAGTTGGAAGAACTGGAGTT 6708
Db 1084 GCACATGACACACACTATGATTTGTTCACTGAGTGTGGAGTTGGAAGAACTGGAGTT 1143
Qy 6709 TTTATTGCTCTGACCATTTAACAACATATATATATATATATATATATATATATATATAT 6768
Db 1144 TTTATTGCTCTGACCATTTAACAACATATATATATATATATATATATATATATATATAT 1203
Qy 6769 GGACTAGTAGCTGAACCTGAGAAAGTGAAGAATGTGATGGTGACAGATCTGGCACAGTAT 6828
Db 1204 GGACTAGTAGCTGAACCTGAGAAAGTGAAGAATGTGATGGTGACAGATCTGGCACAGTAT 1263
Qy 6829 ATCTTTTACACAGTGCAATCTGGATCTCTTATCAAAATAGGGAAGTAAATCAGCCCATC 6888
Db 1264 ATCTTTTACACAGTGCAATCTGGATCTCTTATCAAAATAGGGAAGTAAATCAGCCCATC 1323
Qy 6889 TGTTTGTGTTAATTCAGCACTTCAGAGATGGACTCTTTGGACGCCATGGAAGTGAT 6948
Db 1324 TGTTTGTGTTAATTCAGCACTTCAGAGATGGACTCTTTGGACGCCATGGAAGTGAT 1383
Qy 6949 GTTGAGCTTGAATGGGAAGAACCACTATGTAAATATTCAGACCAAGGATCAAAATGGA 7008
Db 1384 GTTGAGCTTGAATGGGAAGAACCACTATGTAAATATTCAGACCAAGGATCAAAATGGA 1443
Qy 7009 AGAGATTTTAAATCCAGGGGCCAAAGTTA CCCCCTCATTTCTTCGGAATGAAATGTGC 7068
Db 1444 AGAGATTTTAAATCCAGGGGCCAAAGTTA CCCCCTCATTTCTTCGGAATGAAATGTGC 1503
Qy 7069 AACCTTAAGAAATATCTATCTCTCTCCTCACTGCGCTTT 7108
Db 1504 AACCTTAAGAAATATCTATCTCTCTCCTCACTGCGCTTT 1543

RESULT 12

US-10-314-232-10
; Sequence 10, Application US/10314232
; Publication NO. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23

; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1.1
; SEQ ID NO 10
; LENGTH: 2309
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31D
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1)..(2309)
; OTHER INFORMATION: n = unknown nucleotide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (65)..(1030)
US-10-314-232-10

Query Match 15.1%; Score 1071.6; DB 15; Length 2309;
Best Local Similarity 99.4%; Pred. No. 1.3e-244;
Matches 1087; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

Qy 6015 AACGCGGCAATAAGCAAGAAATCCTTCTGCAACATGTTGAAGAGCTTTGCACAACAA 6074
Db 70 AATGAGGCAATAAGCAAGAAATCCTTCTGCAACATGTTGAAGAGCTTTGCACAACAA 129
Qy 6075 CAACCTAAAGTTTCAAGAAGAAATTTTCGGAATTACCAAAATTTCTTCAGGATCTTTCTTC 6134
Db 130 CAACCTAAAGTTTCAAGAAGAAATTTTCGGAATTACCAAAATTTCTTCAGGATCTTTCTTC 189
Qy 6135 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAAAATATAAACCATATA 6194
Db 190 AACTGATGCTGATCTGCTTGGAAATAGAGCAAAAAACCGTTTCCCAAAATATAAACCATATA 248
Qy 6195 TAATAATAATAACAGATTAAGCTGATAGCTGACCTAGTGTTCAGGTTCCGATTTATAT 6254
Db 249 --ATAATAATAACAGATTAAGCTGATAGCTGACCTAGTGTTCAGGTTCCGATTTATAT 306
Qy 6255 TAATCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC 6314
Db 307 TAATCCAGCTATATTTCTGGTTATTTATGTCCTCAATGAATTTATTTGCTACTCAAGTCC 366
Qy 6315 ACTACCAGGAACAGTTGGAGATTTTGGAGATGTTGGGAAAACGAGGCAAAAAACATTT 6374
Db 367 ACTACCAGGAACAGTTGGAGATTTTGGAGATGTTGGGAAAACGAGGCAAAAAACATTT 426
Qy 6375 AGTAATGCTAACACAGTGTTTTGAAGAGGAGCGATCAGATGCCATCAGTATTGGCCAGA 6434
Db 427 AGTAATGCTAACACAGTGTTTTGAAGAGGAGCGATCAGATGCCATCAGTATTGGCCAGA 486
Qy 6435 GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGAATTAACAAAGCTAAATGGAGATGT 6494
Db 487 GGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGAATTAACAAAGCTAAATGGAGATGT 546
Qy 6495 TCAAAATAGATTGGACTATCAGGGATCTGAAAATTCGAAAGGCGATGGGATTTGCATGCTGT 6554
Db 547 TCAAAATAGATTGGACTATCAGGGATCTGAAAATTCGAAAGGCGATGGGATTTGCATGCTGT 606
Qy 6555 TCGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCT 6614
Db 607 TCGACAGTGTAACTTTACTGCTGCGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCT 666
Qy 6615 AATTCACCTTTGTGAAGTTGGTTTCGAGCAAGGAGCGACATGACACACACCTATGATTTGT 6674
Db 667 AATTCACCTTTGTGAAGTTGGTTTCGAGCAAGGAGCGACATGACACACACCTATGATTTGT 726
Qy 6675 TCACCTGCACTGCTGGAGTTGGAAGAACTGGAGTTTATTTGCTCTGGACCATTTAACACA 6734
Db 727 TCACCTGCACTGCTGGAGTTGGAAGAACTGGAGTTTATTTGCTCTGGACCATTTAACACA 786
Qy 6735 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGCTAGCTAGCTAGCTAGCTAGCTA 6794
Db 787 ACATATAAATGACCATGATTTTGTGGATATATATGACTAGCTAGCTAGCTAGCTAGCTAGCTA 846

Qy	6795	AAGAATGTGCATGTGTGCAGAAATCTGGCA	CAGTATATCTTTTAA	CACAGTGCATTTCTGGA	6854
Db	847	AAGAATGTGCATGTGTGCAGAAATCTGGCA	CAGTATATCTTTTAA	CACAGTGCATTTCTGGA	906
Qy	6855	TCTCTTATCAAAATGAAGGAAGTAATCAGCC	ATCTGTGTTTGTAACTATTTCAGCACTTCA	6914	
Db	907	TCTCTTATCAAAATGAAGGAAGTAATCAGCC	ATCTGTGTTTGTAACTATTTCAGCACTTCA	966	
Qy	6915	GAAGATGGAATCTTTTGGACGCCATCGAAG	TGATGTTGAGCTTGAATGGGAAGAAACCCAC	6974	
Db	967	GAAGATGGAATCTTTTGGACGCCATCGAAG	TGATGTTGAGCTTGAATGGGAAGAAACCCAC	1026	
Qy	6975	TATGTAAATATTCAGACCAAGGATACAATT	GGAAGAGATTTTTTAAATCCCGGGGCCAA	7034	
Db	1027	TATGTAAATATTCAGACCAAGGATACAATT	GGAAGAGATTTTTTAAATCCCGGGGCCAA	1086	
Qy	7035	AGTTACCCCTCATCTTCCGAATTGAAATGT	GCAACTTTAAAGAAATATCTATGCTTCT	7094	
Db	1087	AGTTACCCCTCATCTTCCGAATTGAAATGT	GCAACTTTAAAGAAATATCTATGCTTCT	1146	
Qy	7095	CTCACTGTGCCTTT	7108		
Db	1147	CTCACTGTGCCTTT	1160		

RESULT 13

```

US-10-314-232-6
; Sequence 6, Application US/10314232
; Publication No. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO. 6

```

	Query Match	11.1%; Score 789.6; DB 15; Length 2173;
	Best Local Similarity	86.9%; Pred. No. 2.8e-177;
	Matches 951; Conservative	0; Mismatches 4; Indels 139; Gaps 2;
Qy	6015 AACGCGCCCAATAGCAAGAATACTTCCTGCACATGTGGAGAGCTTTGCCAACAA	6074
Dd	70 AATGAGCGCCAATRAGCAAGAAATCCTTCTTGCAACAATGTTGAAGAGCTTTGCCAACAA	129
Qy	6075 CAACCTAAGTTCACAAGAAATTTTCGGAAATTACCAAAATTTCTTCAGCATCTTTCTTC	6134
Dd	130 CAACCTAAGTTTTCAAGAGAATTTTCGGAAATTACCAAAATTTCTTCAGCATCTTTCTTC	189
Qy	6135 AACTGATGCTGATCTGCTTGGTAATGAGCAA AAAACCGTTTCCCACAATAAAAAACATA	6194

Db	190	AAC	TGATGCTGATCTGCTTTGGAAATAGAGCAAAAACCGCTTCCCAACAATAAAACCAT - 248
Qy	6195	TAATAATAA	TAAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGAGTTCGGATTATAT 6254
Db	249	--ATAA	TAAATAACAGAGTAAGCTGATAGCTGACGCTAGTGTTCAGAGTTCGGATTATAT 306
Qy	6255	TAATGCCAGCTA	TATTTCTGGTTTATTTATCTCCAAATGAAATTTATTTGCTACTCAAGGTCC 6314
Db	307	TAATGCCAGCTA	TATTTCTGGTTTATTTATCTCCAAATGAAATTTATTTGCTACTCAAGGTCC 366
Qy	6315	ACTACCA	GGAAACAGTTTGGAGATTTTGGAGATGGTGTGGGAAACCAAGGGCAAAAAACATT 6374
Db	367	ACTACCA	GGAAACAGTTTGGAGATTTTGGAGATGGTGTGGGAAACCAAGGGCAAAAAACATT 426
Qy	6375	AGTAATGCTA	ACACAGTGTTTTGAAGAAGACCGATCAGATGCCATCAGTATTTGGCCAGA 6434
Db	427	AGTAATGCTA	ACACAGTGTTTTGAAGAAGACCGATCAGATGCCATCAGTATTTGGCCAGA 486
Qy	6435	GGACAACA	GCAGCTACTGCTCTTTGGAGATATAGTGATTACAAAGCTAATCAGAGATGT 6494
Db	487	GGACAACA	GCAGCTACTGCTCTTTGGAGATATAGTGATTACAAAGCTAATCAGAGATGT 546
Qy	6495	TCAAATAG	ATTGGATACAGGGATCTGAAAAATTGAAAGCATGGGGATTCGATGACTGT 6554
Db	547	TCAAATAG	ATTGGATACAGGGATCTGAAAAATTGAAAGCATGGGGATTCGATGACTGT 606
Qy	6555	TCGACAGGTAA	CTTTTACGCTGGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCT 6614
Db	607	TCGACAGGTAA	CTTTTACGCTGGCCAGAGCATGGGTTCTTGAGAACAGCGCCCTCT 666
Qy	6615	AAATTCACTTT	GTGAAGTTGGTTTGAGCAACGAGGCAATGACACCACTATGATTGT 6674
Db	667	AAATTCACTTT	GTGAAGTTGGTTTGAGCAACGAGGCAATGACACCACTATGATTGT 726
Qy	6675	TCAC	TGAGTGTGGAGTTGGAAGACTGGAGTTTTTATGCTCTGAGCAATTAACACA 6734
Db	727	TCAC	TGAGTGTGGAGTTGGAAGACTGGAGTTTTTATGCTCTGAGCAATTAACACA 734
Qy	6735	ACATA	TAAATGACCATTGATTTTGTGGATATATATGGACTAGTAGTGAACCTGAGAAGTGA 6794
Db	735	ACATA	TAAATGACCATTGATTTTGTGGATATATATGGACTAGTAGTGAACCTGAGAAGTGA 734
Qy	6795	AAGAA	TGTCATGGTGAGAAATCTGGACAGTATATCTTTTATACAACAGTGCATTTCTGGA 6854
Db	735	-----GGCA	CAGTATATCTTTTATACAACAGTGCATTTCTGGA 770
Qy	6855	TCTCTTAT	CAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTTACGACACTTCA 6914
Db	771	TCTCTTAT	CAAAATAAGGGAAGTAATCAGCCCATCTGTTTGTAACTATTTACGACACTTCA 830
Qy	6915	GAAGATG	GACTCTTTTGGACGCCATGGAAGGTGATTTGAGCTTTGAAATGGGAAGAAACCAAC 6974
Db	831	GAAGATG	GACTCTTTTGGACGCCATGGAAGGTGATTTGAGCTTTGAAATGGGAAGAAACCAAC 890
Qy	6975	TATGTA	AAATTTGACCAAAAGATACAATTGGAAGAGATTTTTTAAATCCACGGGGCAA 7034
Db	891	TATGTA	AAATTTGACCAAAAGGATACAATTGGAAGAGATTTTTTAAATCCACGGGGCAA 950
Qy	7035	AGTTTAC	CCCTCATTTCTTCCGAATTGAAATGTGCAACTTTAAAGAAATATCTATGCTTCT 7094
Db	951	AGTTTAC	CCCTCATTTCTTCCGAATTGAAATGTGCAACTTTAAAGAAATATCTATGCTTCT 1010
Qy	7095	CTCA	TGTGCTTTT 7108
Db	1011	CTCA	TGTGCTTTT 1024

RESULT 14
US-10-314-232-8
; Sequence 8, Application US/10314232
; Publication No. US20030138932A1
; GENERAL INFORMATION:

```
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR fragment
; NAME/KEY: CDS
; LOCATION: (3)..(530)
US-10-314-232-8

Query Match          7.5%; Score 532; DB 15; Length 532;
Best Local Similarity 100.0%; Pred. No. 3.9e-116;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6437 ACAACAAGCCAGTACTGCTTTGGAGATATAGTATCAAAAGCTAATGGAGGATGTTTC 6496
DB 1 ACAACAAGCCAGTACTGCTTTGGAGATATAGTATCAAAAGCTAATGGAGGATGTTTC 60

QY 6497 AAATAGATTGGACTATCAGGATCTGAAATTTGAAAGGATGGGATTCATGACTGTTTC 6556
DB 61 AAATAGATTGGACTATCAGGATCTGAAATTTGAAAGGATGGGATTCATGACTGTTTC 120

QY 6557 GACAGTGTAACCTTACTGCTGGCCAGCATGGGTTCTTGAGAACAGCGGCCCTCTAA 6616
DB 121 GACAGTGTAACCTTACTGCTGGCCAGCATGGGTTCTTGAGAACAGCGGCCCTCTAA 180

QY 6617 TTCACCTTGTGAAGTTGGTTTCGAGCAAGCAGGCGCATGACACCAACCTATGATTGTTTC 6676
DB 181 TTCACCTTGTGAAGTTGGTTTCGAGCAAGCAGGCGCATGACACCAACCTATGATTGTTTC 240

QY 6677 ACTGAGTCTCGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCAATTTAACACAAC 6736
DB 241 ACTGAGTCTCGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCAATTTAACACAAC 300

QY 6737 ATATAAATGACCATGATTTTGTGGATATATATGGAAGTGTGACTGAACTGGAAGTGAAA 6796
DB 301 ATATAAATGACCATGATTTTGTGGATATATATGGAAGTGTGACTGAACTGGAAGTGAAA 360

QY 6797 GAATGTGATGTGCAGAAATCTGGCAGCATATATCTTTTACACAGATGCTATCTTGGATC 6856
DB 361 GAATGTGATGTGCAGAAATCTGGCAGCATATATCTTTTACACAGATGCTATCTTGGATC 420

QY 6857 TCTTATCAAAATAGGGAAGTAATCAAGCCATCTGTTTTGTTTAACATTTACAGCACTTCAGA 6916
DB 421 TCTTATCAAAATAGGGAAGTAATCAAGCCATCTGTTTTGTTTAACATTTACAGCACTTCAGA 480

QY 6917 AGATGAGCTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAGAGA 6968
DB 481 AGATGAGCTCTTTGGACGCCATGGAAGGTGATGTTGAGCTTGAATGGGAGAGA 532
```

RESULT 15

```
US-10-314-232-3
; Sequence 3, Application US/10314232
; Publication NO. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
```

```
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31
; NAME/KEY: CDS
; LOCATION: (1)..(321)
US-10-314-232-3
```

```
Query Match          4.5%; Score 319.4; DB 15; Length 321;
Best Local Similarity 99.7%; Pred. No. 1.7e-65;
Matches 320; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6355 GAAACCAGGGCAAAAACATTAGTAATGCTAACAACAGTGTGTTTTGAAAAAGGACGGATCAGA 6414
DB 1 GAAACCAGGGCAAAAACATTAGTAATGCTAACAACAGTGTGTTTTGAAAAAGGACGGATCAGA 60

QY 6415 TGCCATCAGTATTTGGCCAGAGGACAAACAGCCAGTTACTGTCTTTTGGAGATATAGTGATT 6474
DB 61 TGCCATCAGTATTTGGCCAGAGGACAAACAGCCAGTTACTGTCTTTTGGAGATATAGTGATT 120

QY 6475 ACAAAGCTAATCGAGGAGTGTTCAAATAGATTGGACTATCAGGATCTGAAAAATTCGAAAGG 6534
DB 121 ACAAAGCTAATCGAGGAGTGTTCAAATAGATTGGACTATCAGGATCTGAAAAATTCGAAAGG 180

QY 6535 CATGGGGATTGCATGACTGTTCGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGGTT 6594
DB 181 CATGGGGATTGCATGACTGTTCGACAGTGTAACTTTTACTGCTGGCCAGAGCATGGGGTT 240

QY 6595 CCTGAGAACAGCGCCCTCTTAATTCACCTTTGTGAAGTTGGTTTCGAGCAAGCAGGGCACAT 6654
DB 241 CCTGAGAACAGCGCCCTCTTAATTCACCTTTGTGAAGTTGGTTTCGAGCAAGCAGGGCACAT 300

QY 6655 GACACCACACCTATGATTGTT 6675
DB 301 GACACCACACCTATGATTGTT 321
```

Search completed: July 3, 2005, 10:49:58
Job time : 5493 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 20:48:19 ; Search time 32708 Seconds
(without alignments)
8272.008 Million cell updates/sec

Title: US-10-673-885-1
Perfect score: 7108
Sequence: 1 taattgtgactgtccagaa.....gcttctctactgtgcttt 7108

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2399.2	33.8	3416	3	AK041657	AK041657 Mus muscu
2	625.2	8.8	761	6	CB167019	CB167019 BTU602600
3	491	6.9	561	2	BE668393	BE668393 157119 MA
4	485.2	6.8	538	2	BE751055	BE751055 202768 MA
5	401.8	5.7	517	4	BM750287	BM750287 K-EST0025
6	388.4	5.5	689	2	BB230596	BB230596 BB230596
7	380	5.3	605	2	BB638355	BB638355 BB638355
8	355.4	5.0	905	5	BU222226	BU222226 603106463
9	301.6	4.2	484	8	AQ812654	AQ812654 HS 5266 B
10	277	3.9	771	9	CE385339	CE385339 tigr-gss-
11	230.8	3.2	412	2	AW081039	AW081039 xc39f04.x
12	215	3.0	746	5	BU222294	BU222294 603105362
13	190.8	2.7	629	8	B66492	B66492 CIT-HSP-201
14	190.6	2.7	6600	3	CR749546	CR749546 Homo sapi
15	180.6	2.5	547	8	BH120957	BH120957 RPI-24-2
16	180.6	2.5	713	9	AG323589	AG323589 Mus muscu
17	175.8	2.5	779	7	CF285207	CF285207 AGENCOURT
18	175	2.5	756	9	CC914861	CC914861 t084g17ba
19	169.4	2.4	921	6	CD300952	CD300952 AGENCOURT
20	168.2	2.4	703	3	AQ587839	AQ587839 CITBI-EI-
21	165.4	2.3	6038	3	CR749277	CR749277 Homo sapi
22	161.6	2.3	653	5	BW559375	BW559375 BW559375
23	160.8	2.3	707	5	BW484146	BW484146 BW484146
24	158	2.2	583	7	CK315103	CK315103 SB02041B1

25	155.4	2.2	858	7	CR428355	CR428355
26	154.8	2.2	919	5	BQ710210	BQ710210 AGENCOURT
27	153.8	2.2	707	7	CN790080	CN790080 4124633 B
28	153.8	2.2	779	7	CK777687	CK777687 964818 MA
29	153.4	2.2	744	6	CA512045	CA512045 UI-R-EJO-
30	153.2	2.2	835	6	CD804595	CD804595 UI-M-GVO-
31	151.8	2.1	671	8	AQ580682	AQ580682 RPI-11-4
32	150.8	2.1	694	6	CD353637	CD353637 UI-M-GMO-
33	149.8	2.1	719	9	AG323639	AG323639 Mus muscu
34	148.8	2.1	1191	6	CD504171	CD504171 CDA66-E03
35	146.6	2.1	774	4	BI105833	BI105833 60289253
36	145.6	2.0	859	7	CN325188	CN325188 AGENCOURT
37	144.8	2.0	594	7	CF181137	CF181137 817938 MA
38	144	2.0	732	6	CD350610	CD350610 UI-M-GIO-
39	144	2.0	775	6	CA328425	CA328425 UI-M-FYO-
40	143.4	2.0	664	1	AU206194	AU206194 AU206194
41	143.2	2.0	4286	9	AY417567	AY417567 Homo sapi
42	141.8	2.0	650	1	AU130535	AU130535 AU130535
43	141.6	2.0	701	7	CK807021	CK807021 AGENCOURT
44	141	2.0	3498	3	AK078884	AK078884 Mus muscu
45	140.8	2.0	628	8	AQ243425	AQ243425 HS_2058_A

ALIGNMENTS

RESULT 1	AK041657	AK041657	3416 bp	mrna	linear	HTC 03-APR-2004
LOCUS	AK041657					
DEFINITION	Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:AG30028P16 product:hypothetical Fibronectin type III domain/Fibronectin type III repeat containing protein, full insert sequence.					
ACCESSION	AK041657					
VERSION	AK041657.1 GI:263334640					
KEYWORDS	HTC; CAP trapper.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
REFERENCE	1 Carninci, P. and Hayashizaki, Y.					
AUTHORS	Carninci, P. and Hayashizaki, Y.					
TITLE	High-efficiency full-length cDNA cloning					
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)					
MEDLINE	99279253					
PUBMED	10349636					
REFERENCE	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.					
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.					
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes					
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)					
MEDLINE	20499374					
PUBMED	11042159					
REFERENCE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yamawake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.					
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yamawake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.					
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer					
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)					
MEDLINE	20530913					
PUBMED	11076861					
REFERENCE	4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.					
AUTHORS	Functional annotation of a full-length mouse cDNA collection					
TITLE	Nature 409, 685-690 (2001)					
JOURNAL	Nature 409, 685-690 (2001)					
REFERENCE	5					

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

REFERENCE

6 (bases 1 to 3416)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, S., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/

FEATURES

source

Location/Qualifiers

1..3416

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:A630028F16"

/db_xref="taxon:10090"

/clone="A630028F16"

/tissue_type="thymus"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="3 days neonate"

157..>3414

/notes="unanned protein product; hypothetical Fibronectin

type III domain/Fibronectin type III repeat containing

protein (InterPro|IPR003961, InterPro|IPR003962, evidence:

InterPro)

putative"

/codon_start=1

/protein_id="BAC31021.1"

/db_xref="GI:26334641"

/translation="MDLFFLFLSLIGTSQVDVSGSPDDTVYDITLSSISATTSS

PVSRITATNVKSPGPPVFLAGRVSGAGILLSWNPNNPGRIRIISYVVKYKEVCPMMQ

TAIVRAKPDLSLELLTNLPQTVEIKVAENSAIGVFSDFPFPQAESAPGVV

NLTVEALNYSVNLINWLPQNGKITSFKI SVKHARGSIUVKDVSIKVEDLLSGKLP

ECNENSDSLMSTSPSLSRATPLRTHLSNLRNKLSSVWKPEPISFVYTHLRP

YTYLFEVSAVTEAGIDISTIVRPESVPEGPQNCITGNVTKAFSLISWDPFPAIVT

GFPSVRELYGFTGRIDNSTDLRFVFLTPFTMYDVYVAASAGVSGPKSNLSVF

QDQVYVNNPVTQMTDIAAHLSTYIRLVLPTEHTISVSFTWVGEPPTVLTWRV

OBSPVYVNNPVTQMTDIAAHLSTYIRLVLPTEHTISVSFTWVGEPPTVLTWRV

EQVPSQIQLNYKINSSSILLYWDPPEYNGKIHTYIYAMELDTNRAFQMTWDS

FLITGLKKYTRKMRVAASHTVGSSESLNDLFRVTPDEPSSPDQVKKVTDVDSPE

LSLWSPKPKNGIITAYFYVONADALFNKSTNTITSLDKPFTLYNISIQSYR

LGHQNSLSLVSRTSETVDSAPENITKNTSSSEIETFFLPSPNGIIIOKTYIL

KRSNSHEARTIETTLTIGGLKYTHVIVSVASITLKGEGVRSMPISILTEEDAPD

ORIGIN

Query Match 33.8%; Score 2399.2; DB 3; Length 3416;
Best Local Similarity 84.2%; Pred. No. 0;
Matches 2754; Conservative 0; Mismatches 508; Indels 10; Gaps 4;
QY 1 TAAATGTGTACTTGCAGAGGATCTGTCTTTAAATCAATTAATGCAGGCAACATTTCTCT 60
DB 51 TAAATGTGTACTTGCAGAGGATCTGTCTTTAAATCAATTAATGCAGGCAACATTTCTCT 110
QY 61 CTAGAGCCATCAATGTATCTTACTGGCTG-AAAAATGTAATAAAGATGGATTTCTTAT 119
DB 111 CTAGAGCCATCAATGTATCTTACTGGCTG-AAAAATGTAATAAAGATGGATTTCTTAT 170
QY 120 CAATTTTCTTTTACTTTTATTTGGGATCTTACAGAGCACAGGTTGGTGTTCATATGTCT 179
DB 171 CTTTCTTTCTTTTCTCTCTCTCATTTGGAATTTACAGAGTCACAGGTTGGTGTCTTCT 230
QY 180 TCTGTGTACTAGGTAGGATATTAACCATCTCTTCAATTTCT---ACAACATACACCTCACC 236
DB 231 TGAATGACACTGTGTATGATATCACGCTGTCTTCAATTTCTGCCCAACACCTACAGCTCAC 290
QY 237 TGTACTAGAAATAGTGACACCAAAATGTAACAAACACAGGCTCCAGTCTCTCTAGCCGG 296
DB 291 CGTGTGTAGAACCTGTGGCAACAAATGTGTCAAAACCGGGCCCTCCAGTCTCTCTAGCCGG 350
QY 297 GGAAAGAGTCCGATCTGTCTGGGATTTCTTCTGTCTTGGAAATACACCACTTAATCCAAATGG 356
DB 351 GGAGAGAGTGTGCTCTGCGGGATTTCTCTCTCATGGAATACCCCTCTTAACCCAAATGG 410
QY 357 AAGGATATATCTTCAATGTCAATATTAAGAGTTTGTCCGTGGATGCAACAGATATA 416
DB 411 AAGAATCATATCTGTATGTTGTTAAATATCAAGGAAGTGTCTCCATGGATGACAGACGTA 470
QY 417 TACCAAGTCAGATCAAAAGCAGACAGTCTGGAAGTTCTTCTTAATCTTAATCTCTGG 476
DB 471 TACCGAGTGAGAGGAGGAGCCGACAGTCTGGAAGTTCTTCTCACTAACCTTAAACCTGG 530
QY 477 AACCAATATGAAATTAAGTTTGTCTGTAACCAAGTGTCTGCAATGGAGTGTGTAAGA 536
DB 531 AACCAATATGAAATTAAGTTTGTCTGTAACCAAGTGTCTGCAATGGAGTGTGTAAGA 590
QY 537 TCAATTTCTTCCAAATCTGAGAAAGTCTCCAGAAAGTGTGTAATCTCACAGTTCA 596
DB 591 CCCATTTCTTTTCAAAACCGCAGAAAGTCTCCAGGAAAGTGTGTAATCTCACAGTTCA 650
QY 597 GGCCTACACGCTTCCAGCAGTTAAGTGTGTAATTTGATTTTACCTCGGCAACCAATGGCAA 656
DB 651 GGCCTTCAACTATCTGCAAGTAAATCTGTAATTTGGTATTTTACCTCGGCAACCAATGGTAA 710
QY 657 AATTACACGCTTCAAGATTTAGTGTCAAAACATGCCAGAGTGGGATAGTAGTGAAGATGT 716
DB 711 AATCACGCTTCAAGATTTAGTGTCAAAACATGCCAGAGTGGGATAGTAGTGAAGATGT 770
QY 717 CTCAATCAGAGTAGGAGCAATTTTGACTGGGAAATTTGCCAGATGCAATGCAATGAGATGTA 776
DB 771 CTCAATCAGAGTAGGAGCAATTTTGACTGGGAAATTTGCCAGATGCAATGCAATGAGATGTA 830
QY 777 ATCTTTTATATGAGGTACAGCCAGCCCTTCTCCAAACCTTGTGTAGAGTTACACCTCCATC 836
DB 831 CTCTTTTCTGTGGAGTACCAACAGCCCTTCTCCAAACCTTGTGTAGAGTTACACCTCCATC 890
QY 837 GCGTACCAACATTTCAATCAAGCAGCTTGTGACAGAGATGAGATCAGCTCTGTGTGGAAGA 896
DB 891 TCGTACTACATTTTATCCACACATTTGCGACGGAATAGATCAGCTCTGTGTGGAAGA 950
QY 897 GCGTACCAACATTTCAATCAAGCAGCTTGTGACAGAGATGAGATCAGCTCTGTGTGGAAGA 956
DB 951 GCGTACCAACATTTTATCTTGTGAGTCTCTTCTTGTGAGTCTCTCTTGTGAGTCTCTCTT 1010

QY 957 AGCTGCTCAAACTGAAGCAGGTTATATTGATAGTACGATTTGTGACAAACACCAGAAATCAGT 1016
DB 1011 AGCCGTTACGACGGAAGCAGGGTATATTGACAGTACCAATTTGTTAGGACACCAAGATCAGT 1070
QY 1017 GCCTGAAGGACCAACCAAACTGCGTTAAACAGGCAACATCACAGGAAAGTCTTTTCAAT 1076
DB 1071 GCCTGAAGGACCAACCCAGACTGTATACAGGCAATGTACAGGAAAGCGTTTCAAT 1130
QY 1077 TTTATGGGACCCACCAACTATAGTAAACAGGAAATTTAGTTATAGAGTTGAAATTAATGG 1136
DB 1131 TTTATGGGACCCACCAAGCAATTTGTACAGGAAAGTTTGTATACAGAGTTGAAATTAATGG 1190
QY 1137 ACCATCAGGTGCGATTTTGGATTAACAGCAACAAAGACCTCAAGTTTGCATTTCACTAACCT 1196
DB 1191 GCCAAACAGGAAGATTTTGGATAATAGTACAAAGACCTCCGGTTTGTGTTTACGCATCT 1250
QY 1197 AACACCAATTTACAACTGTATGATGCTTATTTGGGCTGAAACAGAGTCCAGGAGTCCGGC 1256
DB 1251 CAGCCATTTACAAATGTATGATGCTTATTTGTCGAGAAACAGTGTGGGTTGGACC 1310
QY 1257 CAACTCAATTTTCACTATTTCACTCCACAGATTTCCAGGGGCAAGTGTGTTGATTTACA 1316
DB 1311 CAACTCAATTTTCACTATTTCACTCCACAGATTTCCAGGAGCAAGTGTGTTGATTTACA 1370
QY 1317 ACTTGACAGGTTAGAAATCCACGCAAGTAAAGATTAATCTTGGAGAAACCCAGCAACCAAA 1376
DB 1371 GATCGTGGAGGTAGAAAGCCACAGAAATTAAGATTTCTTGGAGGAAACCTCGCAACCAAA 1430
QY 1377 TGGAAATTTAACCAATACCGAGTGAAGTGTAGTTTCCAGACACAGGAATTAATTTGGA 1436
DB 1431 TGGAAATTCATCAGCCAGTACCGTGAAGTGTAGTCTCAGTGTAGAGTCAAGGATCAATTTGGA 1490
QY 1437 AAATACCTTTGCTCACTGGAAATTAATGAGTATATAAATGACCCCATGGCTCCAGAAATTTG 1496
DB 1491 AAACACTCTGCTCACAGGACAGATGAGTATATAACAAACCCATGACTCCAGAAATAT 1550
QY 1497 GAACTAGTAGAGCCAAATGGTAGGATTAATAGAGGTTTCCAGAGAGATGTGCTCACT 1556
DB 1551 GAACTAGTAGAGCCAAATGATAGGATTTATAGGCTTATAGAGGCTCAGGAGAGATGCTTCACT 1610
QY 1557 TCATCACTTGTCTACATTTATATAAAGCCATCCAGATATAAATCTTCTCCAGGAA 1616
DB 1611 GCATCCCTCGCTTCGTTTCATATAACACAGCCATCCA---CAGACTTTTCTCCAGGAC 1667
QY 1617 TAGAGCTGAAGACCAAGACTTCAACAGTTGTA---ACTAAGGAATCAGTATATTAATCTGA 1673
DB 1668 AAGAGTTGAAGACCAAGATCACAGTTGATGAGGCAACAGGAAATCAGTATATGACTGA 1727
QY 1674 CATTTGACGTGAACAGCTGTCTTATGTTATCAGGAGACTTTGTACTTTCACTGAGCACAT 1733
DB 1728 CATTTGCTGTAACACTGTCTTATGTTATCAGGCGCTTGTACCTTCACTGAGCACAC 1787
QY 1734 GATTAGTATCTGCTTTTCAACATATGAGGAGAGGACCAACCAAGTCTCAGTGTAG 1793
DB 1788 GATTAGCTATCTGCTTTTCACTGTCTATGCGGAGGACCAACCAAGTCTCACTGTTAG 1847
QY 1794 GACAGCTGACAGTGCAGCTCCATTAATAATTAATAACTATAAATAATTAATAGTTCTTC 1853
DB 1848 GACAGCTGACAGTGCAGCTCCATTAATAATTAATAACTATAAATAATTAATAGTTCTTC 1907
QY 1854 ATCTATTTTGTATATTGGGATCTCCAGAAATATCCCAATGGAAAAATAACTCACTATAC 1913
DB 1908 ATCTATTTTGTATATTGGATCTCCAGAAATATCCCAATGGAAAAATAACTCACTATAC 1967
QY 1914 GATTATGCAATGGAAATTTGATACAAACAGAGCAATTTCCAGATTAATCAACATAGATAACAG 1973
DB 1968 AATCTATGCAATGGAAATTTGATACAAACAGAGCTTTCCAGATGACTACTGTAGATAACAG 2027
QY 1974 CTTTCTCATACAGGGTTAAGAAATACACAAATACAAATGAGAGTGCAGCTCAAC 2033
DB 2028 CTTTCTCATACAGGGTTAAGAAATATACAAATATTAATAAGTAAATGAGGTTGCTGCCTCAAC 2087

QY 2034 CCACGATGAGAAAGTTCTTCTGTAAGAAATGACATCTTTGTGAGAACTTTCAGAAGA 2093
DB 2088 TCATGTTGGGAAAGTTCTCTGTGAGAAATGACCTCTTTGTGAGAACTTTCAGAAGA 2147
QY 2094 TGAACCGGAATCATCACCTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAAATAAG 2153
DB 2148 TGAACCGGAATCATCACCTCAAGATGTCGAAGTAAATGATGTTTACCGCAGATGAAATAAG 2207
QY 2154 GTTGAAGTGGTCCACCCGAAAGCCCAATGGGATCATTTGTTGTTATGAAAGTGTATTA 2213
DB 2208 CTTGACCTGGTCCCACTGAGAAACCTAATGGGATCATTTGTTGTTTACCGCAGATGTTT 2267
QY 2214 TAAATATATAGTACTTTTATATATGAGAAACATCAACCAACAGACATATATTAAGAA 2273
DB 2268 CCAAAATGAGATGCTTTGTTGTAAGAAACCTCCACCAAGAAATATTAAGTGA 2327
QY 2274 CTTAAGACCTTCACACCTCTATAACATTTCTGTAAGTCTTACACAGATTTTGGTCAATGG 2333
DB 2328 CTTGAAACCTTACACACTCTATATATTTCTATCCAGTATACACAGACTTGGTCAATGG 2387
QY 2334 CAATCAGGTATCTTCTTTACTCTCTGTAAAGACTTCGGAGACTGTGCTGATAGTGCAAC 2393
DB 2388 CAATCAGTCACTCTCAATTTCTTGTAAAGACTTCAGAGACTGTACCTGATAGTGCGCC 2447
QY 2394 AGAAATATACCTTACAAAAATTTTCTTCTGAGAGATTTGAGTATCATTTCTTCCCC 2453
DB 2448 AGAAATATTTACTTACAAAAATTTTCTTCTGAGAGATTTGAGATTTTCTTCTTCTTCC 2507
QY 2454 AAGTAGTCCCAATGGAATCATATAAAAAATATCAATTTATCTCAAGAGAGTAAATGGA 2513
DB 2508 AAGGAGTCCCAATGGAATCATATAAAAAATATCAATTTATCTCAAGAGAGTAAATGGA 2567
QY 2514 TGAGGAAAGAACTATAAAATPACAACTCTTTTAAACCCAAAAATTTAAAGTACTGAAGAA 2573
DB 2568 CGAGGCGAGAACTATAGAAACAACTCTCTCTCACTATCGGAGGACTGAAGAAATGA 2627
QY 2574 TACCCATATATATCATTTGAGGTGTCTGTAGTACACTGAAGGTAAGGAGTTCCGAGTGC 2633
DB 2628 TACCCATATATGTTGAGGTGTCTGCAAGTACACTGAAGGTAAGGAGTTCCGAGTAT 2687
QY 2634 TCCCAATAGTATACCTGAGGAGAGAGTCTCTGATTTCTCCCTCAAGACTTCTCTGT 2693
DB 2688 GCAATTAAGCACTGACTGAGGAGAGAGTCTGATTTCTCTCTCTCAGAACTTCTCTGT 2747
QY 2694 AAAACAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2753
DB 2748 GAAACAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2807
QY 2754 AATATCTCTTATATACAGATTTTATGCTGGAATAGATCATCATTAATAAATATTAATGT 2813
DB 2808 AATATCTCTTATATACAGATTTTATGCTGGAATAGATCATCATTAATAAATATTAATGT 2867
QY 2814 CACTGAAACATCATTTGAGGTATCAGATTTGGAATATATGTTGAATACAGTCTTATGT 2873
DB 2868 AATGAGATCATTTGAGGTATCAGATTTGGAATATATGTTGAATACAGTCTTATGT 2927
QY 2874 AACAGCTAGCACAGATTTGTTGATGGGAAACAGGAGCAATATCATTTAGCTTTTCAAA 2933
DB 2928 AACAGCTAGCACAGATTTGTTGATGGGAAACAGGAGCAATATCATTTAGCTTTTCAAA 2987
QY 2934 ACCAGAGGAGCACCAAGCGATCTCTCCAAAGATGTTTATATGCAAACTCAGTCTTTC 2993
DB 2988 ACCAGAGGAGCACCAAGCGATCTCTCCAAAGATGTTTATGCAAACTCAGTCTTTC 3047
QY 2994 ATCAATATTTCTTTCTGAGACCTCTTCAAAACCTAATGGATTTGATTAATTTACTC 3053
DB 3048 ATCGATAATTTCTTTCTGAGCTCTCTCTGTAACCTAATGGAAATCATCAATATTTATTC 3107
QY 3054 TGTATTTATACAGAAATACCTTCAAGTACTTTTATGAGAAATTTTACACTCCATGAACTAAC 3113
DB 3108 TGTATTTATACAGAAATACCTTCAAGTACTTTTATGAGAAATTTTACACTCCATGAACTAAC 3167
QY 3114 CAATGACTTTGCAATATGACTGTATCCAAATTTATAGATAAATGCAAAATATTCAGCTA 3173

Db 3168 CCAGGACGCTGGCAATGTGACAGTGTCTCAAGAACTCTATAAACTGGCAGTATTCAGCTA 3227
Qy 3174 CTATACATTTGGTTTAAACAGCAAGTACTTTCAGTTGGAAATGGGAATAAAAGCAGTACAT 3233
Db 3228 CTATACATTTGGCTTAAACAGCAAGTACTTTCAGTTGGAAATGGGAATAAAAGCAGTATGT 3287
Qy 3234 CATTGAAGTATACACAGATCAAGACATACCTG 3265
Db 3288 AATTCTATGCTACACAGACCAAGACATTTTCAG 3319

RESULT 2
CB167019/c 761 bp mRNA linear EST 30-JAN-2003
LOCUS BRU602600590.R1 CSEQFL04 small intestine Bos taurus cDNA, mRNA
DEFINITION CB167019
ACCESSION CB167019
VERSION CB167019.1 GI:28153144
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

REFERENCE 1 (bases 1 to 761)
AUTHORS Adelson,D.L. and Gill,C.A.
TITLE Bovine ESTs (Adelson and Gill)
JOURNAL Unpublished (2003)
COMMENT Contact: David L. Adelson
Animal Breeding and Genetics

Texas A&M University
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
USA
Tel: 9798452616
Fax: 9798456970
Email: david.adelson@tamu.edu.

FEATURES
source Location/Qualifiers

1..761
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/cissue type="small intestine"
/clone lib="CSEQFL04 small intestine"
/note="Organ: small intestine; Vector: pBluescript SK+;
Site 1: NotI; Site 2: EcoRI; sequence 5' of the insert
(5'-NNN...NNNinsert)
GCGAATTGGAGCTCCACCGCGTGGCGCGCGCTCGAG. Sequence 3' of
the inserts (AAGAATTCGATATCAAGCTTATCGATACGTCGACCTCGAG.
non-normalized library, sequenced 3' with M13R primer."

ORIGIN

Query Match 8.8%; Score 625.2; DB 6; Length 761;
Best Local Similarity 90.0%; Pred. No. 3e-137;
Matches 682; Conservative 0; Mismatches 73; Indels 3; Gaps 1;
Qy 2767 TACACAGTTTATGCTCGGAATAGATCATCATTTAAAAACTTAAATGCTCACTGAAACATCA 2826
Db 761 TACACAGTTTATGCTCGGCAGACATCATCTTGGAAAACCTGTTAATGTAATCACTGAAACATCA 702
Qy 2827 TTGAGCTTATCAGATTTGGATTATAATGTTGAATACAGTGTCTTATGTAACAGCTAGCACCC 2886
Db 701 TTAGATTTTTCAGATTTGGACCACAAATGTTGAATATATGCTTATGTTAAGCTAGTATACC 642
Qy 2887 AGATTTTGGTATGGGAAAAACAGGAACAATATCATTTAGCTTTTCAAAACACAGAGGGAGCA 2946
Db 641 AGATTTTGGTATGGGAAAAAGAGAAGCAATATCATCACTTCGGAACACCCGAGGGAGCA 582
Qy 2947 CCAGCGATCTCCCAAGATGTTTATATGCAAACTCAGTCTTTCATCAATAATTCCT 3006
Db 581 CCAAGTGATCTCCCAAGATGTTTCAATATTCAGATTTCAAACTCAGTCTTTCATCAATAATTCCT 522
Qy 3007 TTCGAGCACCTCTCTTCAAAACCTAATGGGATTTATACAATATTTACTCTGTTTATTACAGA 3066

Db 521 TTTTGGACACCTCTCTTCAAAACCTAATGGGATTTATACAATATTTCTGTTTATTACAGA 462
Qy 3067 AATACTTCAGGTACTTTTATGAGAAATTTTACACTCCATGAACCAATGACTTTGAC 3126
Db 461 AATACTTCAGGTACTTTTATGAGAAATTTTACACTCCATGAAGTA---AATGACTTTGAC 405
Qy 3127 AATATGACTGTATCCACAATTTATAGATAAACTGACAATATTCAGCTTACTATATCATTTTGG 3186
Db 404 AATATGACCGTATCTGCAATATATAGATAAACTGGCAATATTCAGCTACTATATCATTTTGG 345
Qy 3187 TTAACAGCAAGTACTTTCAGTTTGGAAATGGGAATAAAAGCAGTGAACATCAATGAAGTATAC 3246
Db 344 GTAAACAGCAAGTACTTTCAGTTTGGAAATGGGAATAAAAGCAGTGGCAGTATTCACGTATAT 285
Qy 3247 ACAGATCAAGACATACCTGAAGGTTTGTGGAAACCTGACTTACGAATCCATTTTCGTCA 3306
Db 284 ACAGATCAAGACATACCTGAAGGAGTTGTGGGAAACCTGACTTACGAGTCCATTTTCGTCA 225
Qy 3307 ACTGCAATAAATGTAAGCTGGGTCCACCGGTCCACCAACCGGTCTAGTCTTCTACTAT 3366
Db 224 ACTGAGTAAATGTCTAGCTGGGGCCCTCCGTCTCAACCAATGGTCTAGTCTTCTACTAT 165
Qy 3367 GTTCTCACTGATCTTACAGCAGACTCTCCGCCATGTGAGACCACTCTTGTTCATATGAG 3426
Db 164 GTTCTCTGAGCTCAGCAGCAGACTCGCGCCACGCTGAGACCACCTCTAGTATTAGTATGAG 105
Qy 3427 AGAAGCATATATTTGATTAATCTGGAAAATACACTGATATATATATAAAATTAATCTCCA 3486
Db 104 AGAAGCATGTATTTGATAATCTGGACAAATATCTGACTATATCTTTAAAGTCACTCCA 45
Qy 3487 TCACAGAAAAGGGATTCTCTGATCACTATCTGCCCCA 3524
Db 44 TCACAGAAAAGGGATTCTCTGACGCATATACGGCCCCA 7

RESULT 3

BE668393 561 bp mRNA linear EST 25-APR-2001
LOCUS 157119 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE668393
ACCESSION BE668393
VERSION BE668393.1 GI:10028984
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 561)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnkruug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-Mckown,C.G., Perteu,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

TITLE

Genome Res. 11 (4), 626-630 (2001)
21180013
11282978
JOURNAL Contact: Smith TPL
MEDLINE USDA, ARS, US Meat Animal Research Center
PUBMED PO Box 166, Clay Center, NE 68933-0166, USA
COMMENT Tel: 402 762 4366
Fax: 402 762 4390

Email: smth@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 76 row: I column: 2
Seq primer: ATTTAGGTGACACTATAG.

```
FEATURES
source      Location/Qualifiers
1..561
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 4BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."

ORIGIN
Query Match      6.9%; Score 491; DB 2; Length 561;
Best Local Similarity 93.3%; Pred. No. 2.4e-105;
Matches 525; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

QY 6163 GCAAAAACCGTTTCCCAACATATAAACCATATAATAATACACAGTAAAGCTGATA 6222
DB 1 GCAAAAACCGTTTCCAAACATATAAACCAT---ATAACAAATACAGAGTAAAGCTGATA 57

QY 6223 GCTGAGCTAGTGTTCAGGTTCCGATTATATTAATGCCAGCTATATTTCTGTTATTTA 6282
DB 58 GCTGATGCTAGTATTCAGGATCAGATTATTAATGCCAGCTATGTTCTGGCTACTTA 117

QY 6283 TGTCCTAAATGAATTTATGCTACTCAAGTCCACTACAGGAACAGTTGGAGATTTTGG 6342
DB 118 TGTCCTAAATGAATTTATGCTACTCAAGTCCACTACAGGAACAGTTGGAGCTTTTGG 177

QY 6343 AGAATGGTGGGAACACCGGCAAAACATTAATAGTAAATGCTAACACAGTTTGA 6402
DB 178 AGAATGGTGGGAACACCGGCAAAACATTAATAGTAAATGCTAACACAGTTTGA 237

QY 6403 GGACCGATCAGATGCATAGTATGCGCAGAGCAACAAAGCCAGTTACTGCTTTGGA 6462
DB 238 GGACCGATCAGATGCATCAATATTTGGCAGAGCAACAAACAGTTACTGCTTTGGA 297

QY 6463 GATATAGTATTAACAAGCTAATGGAGGATGTTCAATAGATGGAATGGAATGGA 6522
DB 298 GATATAGTATTAACAAGCTAATGGAGGATGTTCAATAGATGGAATGGAATGGA 357

QY 6523 AAAATTTGAAGGACCGGATTCATGCTGAGTGTTCGACAGTGAATTTACTGCTGCCA 6582
DB 358 AAAATTTGAAGGACCGGATTCATGCTGAGTGTTCGACAGTGAATTTACTGCTGCC 417

QY 6583 GACATGGGTTCTGAGAACACGCGCCCTCTAATTCATCTTGTGAAGTTGGTTCCAGCA 6642
DB 418 GAACACGGGTTCTGAGAACACGCGCCCTCTAATTCATCTTGTGAAGTTGGTTCCAGCA 477

QY 6643 AGCAGGCAATGACACCACTATGATTTGTTCACTGCAAGTCTGGAGTTGGAAGAACT 6702
DB 478 AGCAGGCAATGACACCACTATGATTTGTTCACTGCAAGTCTGGAGTTGGAAGAACT 537

QY 6703 GGAGTTTATTTGCTCTGGACCA 6725
DB 538 GGAGTTTATTTGCTCTGGACCA 560

RESULT 4
BE751055 538 bp mRNA linear EST 25-APR-2001
LOCUS 202768 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE751055
ACCESSION BE751055
VERSION BE751055.1 GI:10165047
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 538)
AUTHORS Smith,T.P.L., Grose,W.M., Preking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,

Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-Mckown,C.G., Perte,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
JOURNAL MEDLINE
PUBMED 11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGAAACACGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 42 row: H column: 16
Seq primer: ATTAGTGCACATATAG.
FEATURES
Location/Qualifiers
1..538
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="MARC 2BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN
Query Match      6.8%; Score 485.2; DB 2; Length 538;
Best Local Similarity 93.9%; Pred. No. 5.8e-104;
Matches 505; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 6419 ATCAGTATGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTATACAA 6478
DB 1 ATCAATATTTGGCCAGAGGACAAACAGCCAGTTACTGCTTTGGAGATATAGTATACAA 60

QY 6479 AGCTAATGGAGGATGTTCAATAGATTGGAGTATCAGGATCTGAAATTTGAAAGGCATG 6538
DB 61 AACTAATGGAGGATGTTCAATAGATTGGAGTATCAGGATCTGAAATTTGAAAGGCATG 120

QY 6539 GGGATTGCATGATCTGTCACAGTGAATTTACTGCTGGCCAGAGCATGGGGTTCCCTG 6598
DB 121 GGGATTGCATGATCTGTCACAGTGAATTTACTGCTGGCCAGAGCATGGGGTTCCCTG 180

QY 6599 AGAACAGCCGCTCTAATTTCACTTTGTAAGTTGTTGAGCAACGAGGCACATGACA 6658
DB 181 AGAACAGCCGCTCTCTAGTCCACTTTGTGAAGTGTTCGAGCAACGAGGCACATGATA 240

QY 6659 CCACACTATGATTTGTTCACTGAGTGTGGAGTTGGAAGAACTGGAGTTTTTATGCTC 6718
DB 241 CCACACCCATGATTTGTTCACTGAGTGTGGAGTTGGAAGAACTGGAGTTTTTATGCTC 300

QY 6719 TGGACCATTTAACACAACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAG 6778
DB 301 TGGACCATTTAACACAACATATAAATGACCATGATTTTGTGGATATATATGACTAGTAG 360

QY 6779 CTGCACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGATCTGGCAGATATCTTTTAC 6838
DB 361 CTGCACTGAGAGTGAAGAAATGTCATGTCAGATCTGGCAGATCTGGCAGATATCTTTTAC 420

QY 6839 ACCAGTGCATCTCGATCTCTTATCAATAAGGAGTAATCAGCCCATCTGTTTGTGTTA 6898
DB 421 ATCAGTGCATCTCGATCTCTTATCAAGTAAGGAAATTAATCAGCTCATCTGTTTGTGTTA 480

QY 6899 ACTATTGACACTTCAGAAAGATGGACTCTTTGGACCCATGGAAGGTGATGTTGAGCT 6956
```

```
Db 481 ACTATTGACACTTCAGAAGATGAGTCTTTGGATGCCATGGAAGTGATGTTGAGCT 538
|||||
RESULT 5
BM750287 517 bp mRNA linear EST 04-MAR-2002
LOCUS K-EST0025803 S5SNU484 Homo sapiens cDNA clone S5SNU484-12-C05 5',
DEFINITION mRNA sequence.
ACCESSION BM750287
VERSION BM750287.1 GI:19079905
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 517)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 12 row: C column: 05
High quality sequence stop: 517.
FEATURES
Location/Qualifiers
1..517
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S5SNU484-12-C05"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/lab_host="SNU-484"
/lab_host="Top10F"
/clone_lib="S5SNU484"
/note="Organ: Stomach; Vector: pTZ19RPI; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was decapped with tabacco
acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
including EcoRI site by treatment of T4 RNA ligase. The
first strand cDNA was synthesized from oligo dt-selected
mRNA by priming with dt-tailed vector. The dt-tailed
vector was adjusted to have about 60nt. The cDNA vector
was circularized with E. coli DNA ligase after digestion
of EcoRI which site is also included in vector. An RNA
strand converted to a DNA strand by Okayama-Berg method.
The obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation
method."
ORIGIN
Query Match 5.7%; Score 401.8; DB 4; Length 517;
Best Local Similarity 99.5%; Pred. NO. 4e-84;
Matches 403; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3718 ATATTGAGAGCTTTCACCATTTACATTATATAGCTTTTTCCTGCCGCAAGACTAGA 3777
DB 3 ATATAGGAAGAGCTTTCACCATTTACATTATATAGCTTTTTCCTGCCGCAAGACTAGA 62
QY 3778 AAAGGACTTGCTTCCAGTATTTCTTTCTTTTACACAGATCAGTCAGTCGCGTTAGCA 3837
DB 63 AAAGGACTTGCTTCCAGTATTTCTTTCTTTTACACAGATCAGTCAGTCGCGTTAGCA 122
QY 3838 CCTCCACAAATTTGACTTTTAATCAACTGCTACTTCAGACTTTGATGGCTGGAATGAGC 3897
DB 123 CCTCCACAAATTTGACTTTTAATCAACTGCTACTTCAGACTTTGATGGCTGGAATGAGC 182
```

```
QY 3898 CCAAGTCTCTTCCAGGTGGTATTGTTAAAGTATATAGTTTTAAATTCATGAACATGAA 3957
DB 183 CCAAGTCTCTTCCAGGTGGTATTGTTAAAGTATATAGTTTTAAATTCATGAACATGAA 242
QY 3958 ACTGACACTATATATTATAAGAAATATATATCAGGATTATAAATCGAAGCCAACTTTGTGGA 4017
DB 243 ACTGACACTATATATTATAAGAAATATATATCAGGATTATAAATCGAAGCCAACTTTGTGGA 302
QY 4018 CTGGAACCACTGAGCAGCCTACTCTATCCGTGTATCTGCTTCACCAAACTGCGAATGCG 4077
DB 303 CTGGAACCACTGAGCAGCCTACTCTATCCGTGTATCTGCTTCACCAAACTGCGAATGCG 362
QY 4078 AATCAATTTAGTAATGTAGTAAATTCACAAACCCCAAGATCAGTT 4122
DB 363 AATCAATTTAGTAATGTAGTAAATTCACAAACCCCAAGATCAGTT 407
689 bp mRNA linear EST 23-OCT-2001
BB230596 RIKEN full-length enriched, 3 days neonate thymus Mus
musculus cDNA clone A630028F16 3', mRNA sequence.
BB230596
ACCESSION BB230596
VERSION BB230596
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 689)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,K., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tanigawa,A., Takahashi,F.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 4, 2000 this sequence version replaced gi:8910509.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,T.,
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genomic Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
```

Db 482 ATAAACTGCGCAGTATTCAGCTACTATACATTTTGGCTAAACAGCAAGTACTTTAGTTGGAA 541

Qy 3212 ATGGGAATAAAGCAGTCAGTCACATCATTTGAAGTATACACAGATCAAGACATACCTG 3265

Db 542 ATGGGAATAAAGCAGTCAGTGAATTCATGTCTACAGACCAAGACATTTTCAAG 595

1. .689

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630028F16"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate
thymus"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

```

ORIGIN

[illegible]

```
e mouse tissues.
Location/Qualifiers
1. 605
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630028F16"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length
thymus"
```

BB638355 605 bp mRNA linear EST 26-OCT-2001
 BB638355 RIKEN full-length enriched, 3 days neonate thymus Mus
 musculus cDNA clone A63028F16 5', mRNA sequence.
 BB638355
 BB638355.1 GI:16474002
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 605)
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,
 Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,
 Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,
 Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Yoko Estate (Arakawa, T., et al. 2001)
 Unpublished (2001)
 Contact: Yoshihide Havaashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Suganara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
Wachi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Suganara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.
Location/Qualifiers
1. .605
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630028F16"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate thymus"

Query Match 3.9%; Score 277; DB 9; Length 771;
Best Local Similarity 88.3%; Pred. No. 2.2e-54;
Matches 301; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4389 AGTTCACAGTGTCCCAAAATATTGCTTTTCTGATGTTCACTCACTAGTGCACATTT 4448
DB 38 AGTTCACAGTGTCCCAAAATATTGCTTTTCTGATGTTCACTCACTAGTGCACATTT 97

QY 4449 GACATGGATGAAGCTGACACTATCTCGGCTACTTTCAAAATTAACAAATTTACCACTCA 4508
DB 98 AACATGGATGAGACCTGACACTATTTTGGCTACTTTCCAAAATTAACAAATTAACCACTCA 157

QY 4509 ACTTCGTGCTCAAAAATGCAAGAATGGCAATCCGAAGAATGTGTTGAATATCAAAAAT 4568
DB 158 ACTCGGGCTCAGAAATGCTGAAATGGGATCTGAAGAATGTGTTGAATATCAAAAAT 217

QY 4569 TCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGAATTAAGAAATTTAG 4628
DB 218 TCAATACCTTTATGAAGCTGACCTAACTGAGAGACAGTATGGAATTAAGAAATTTAG 277

QY 4629 ATGTATAGATTCGAAGTGGCTGCCAGACCAATGCTGCTATGGCAATGCTTCAAACTG 4688
DB 278 ATGTATAGATTCGAAGTGGCTGCCAGACCAATGCTGCTATGGCAATGCTTCAAACTG 337

QY 4689 GATTTCTACAAAATCTCGCTGGCCCTCCAGATGCTCCTC 4729
DB 338 GATATCTACAAAACCTCTCCTGCTGTAAGTATTAATCTTC 378

RESULT 11
AW081039/c
LOCUS
DEFINITION
AW081039
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW081039 412 bp mRNA linear EST 14-OCT-1999
xc39f04.x1 NCI CGAP Co20 Homo sapiens cDNA clone IMAGE:2586655 3'
similar to TR:088488 088488 GLOMERULAR MESANGIAL CELL RECEPTOR
PROTEIN-TYROSINE PHOSPHATASE PRECURSOR ;, mRNA sequence.

AW081039
AW081039.1 GI:6036191
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 412)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 379.
Location/Qualifiers
1. .412
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2586655"
/tissue_type="moderately differentiated adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Co20"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Normalized to Cot 500. Average insert size 1.1kb.
Normalized version of NCI_CGAP_Co18. Library constructed

by Life Technologies."

ORIGIN
Query Match 3.2%; Score 230.8; DB 2; Length 412;
Best Local Similarity 99.1%; Pred. No. 1.9e-43;
Matches 232; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 264 AACAAAACCCAGGCGCTCCAGTCTTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTC 323
DB 394 AATAGAACCGAGGCGCTCCAGTCTTCTAGCCGGGAAAGAGTCGGATCTGCTGGGATTC 335

QY 324 TCTGTCTTGGAAATACACCACTTAATCCAAATCGAAGGATTAATCTTACATTTGTCAATA 383
DB 334 TCTGTCTTGGAAATACACCACTTAATCCAAATCGAAGGATTAATCTTACATTTGTCAATA 275

QY 384 TAAGGAAGTTTTCGCTGGATGCAACACAGTATATACAAAGTCAGATCAAAAGCCAGACAG 443
DB 274 TAAGGAAGTTTTCGCTGGATGCAACACAGTATATACAAAGTCAGATCAAAAGCCAGACAG 215

QY 444 TCTGGAAGTTTCTTCTTACTTAATCTTAACTCTGGAACACATATGAAATTTAAGGT 497
DB 214 TCTGGAAGTTTCTTCTTACTTAATCTTAACTCTGGAACACATATGAAATTTAAGGT 161

RESULT 12
BU222294
LOCUS
DEFINITION
BU222294
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

BU222294 746 bp mRNA linear EST 25-NOV-2002
603105362F1 CSEQCHN04 Gallus gallus cDNA clone CHEST43116 5', mRNA
sequence.
BU222294
BU222294.1 GI:25411515
EST.
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 746)
Boardman, P.E.; Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .746
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hixex"
/db_xref="taxon:9031"
/clone="CHEST43116"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN04"
/note="Organ: whole embryo; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2


```

/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/notes="protein tyrosine phosphatase, receptor type B
precursor, differentially spliced"
1. .6600
/gene="DKFp686H15164"
27. .2786
/gene="DKFp686H15164"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH18347.1"
/db_xref="GI:51476765"
translacion="MEAEFYMVILTCLIPRNSGFOIVHVOQOCLFKNEKVVVGSCL
RIQOQWMTNEKLLHVKSALNSRSPRSAILDCSOAPWTCYDQSGFL
EVENASLFLOKQGRVVKARKYLHSNMKIDYKNEKLVNLSLCKRAGLGAESVR
STRTAPPOLITFNAPDGLFLRLNTEAFIRNAENYSQNSRSHPNLHMTGIT
DYSWLLSTPSSSTTEETGLAEPERCNFTLAESKASHSVSIQWRIILGSPCNFLIY
SSDTLGAALCPTFRIDNTTYGCLQDLQAGTTFNFKIISLDEERTVTLQDPLPARF
GVSKEKTSLSLHVMTPSSGKVSVEVGFQDENNKIQGVOIQESTSWNEVTFNLT
AGSKYNIATVSGGKSPSVVTNGSTVSPVKDIGSTKANSLIISHSGSGNVERV
RLMDXGILVGGVVDKRTSYAFHGLTPGYLYNLTWTEAAGLQNRWKLVRTPAM
EVSNLKXNDGSLTSUKVQRPNGVDNISITLSHKGTIKESRLAPWITETHREL
VGRLYQVTVSCVSGELSAQMAVGRTPDKVANLEANNGMRSLVSWSPAGDWE
QYRIILFNDVVLNLTGVKEETQYVMDDTGLVPGQYEVIEVSGNLKNSRQQR
TVPLAVQLVRKHANETSLSIMWOTPAEWKVIISLADRLLLIHKSLKDKAETFF
TDLVPCRKYMAVTSISGDLKNSSVKGTTPAQVTDLDHVNQGMTSILFTNWTQAG
DVEFQVLLIHENVIKNESISSETSRYSFHSLSKGSLSYVVTVVSGGISRQVVE
GRTVPDKVQGVSVNSARSADYLRVSWHATGDFHYEVTIRKNNPFIQTKAPPSQKTN
VYLFs"

```

ORIGIN

Query Match		2.7%; Score 190.6; DB 3; Length 6600;
Best Local Similarity		54.6%; Pred. No. 1.4e-33;
Matches		476; Conservative 0; Mismatches 374; Indels 21; Gaps 4;
QY	6005	ACGAGAGATAACCGCGCCAAATAAGCAAGAAATCCCTTCCTGCAACATGTTGAAGAGCTTT 6064
Db	5442	ACCGNAAACTTCTTGTCCATAAATAATAATACAGTTTGAAGGCGATTTTCATGAGCTAC 5501
QY	6065	GCACAAACAAACCTAAAGTTTCAAGAGAATTTTCGGAATTAACAAAATTTCTTCAGG 6124
Db	5502	AGGCTGGCTCCAACACTTCTATCCAGGAATACGAGGAGTTAAAGACGCTGGGCGGAA 5561
QY	6125	ATCTTCTTCACTGATGCTGATCTGCTTGGATAGACCAAAACCGTTTCCCAACA 6184
Db	5562	ACCAGTCATGTGACATTCGACTCTTCGCGGAGAATAGAGGGAATAATCGATACAATA 5621
QY	6185	TAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCAGGTT 6244
Db	5622	TATTGCGCT---ATGATGCCAGCGAGTGAGCTTCCAAATGATGATGATCTTGTGT 5678
QY	6245	CGGATTATTAATGCCAGCTATATTCTGGTTATTATGTCCAAATGAATTTATTGCTA 6304
Db	5679	CTGACTACATCAATGCAGCTACATCCCTGGCAACAACCTTCAGAGAGAATACTTGTC 5738
QY	6305	CTCAGGTCCATACAGAGAACGTTGGAGATTTTGGAGATCGTGTGGAAACACAGG 6364
Db	5739	CTCAGGAGCGCTTCTCGGCACCAAGGATGACTTCTGGAATAATGGTGTGGAAACAAACG 5798
QY	6365	CAAAAACATTAGTAATGCTAACACAGTGTTTGAAAAAGAGCGATCAGATCGATCACT 6424
Db	5799	TTCAACAATCGTCATGGTGACCCAGTGTGTGAGAGGCGCGAGTAAGATGTGACCAT 5858
QY	6425	ATTGGCCAGAGGACAAACAGCCAGTTACTGTCTTTGGAGATATAGTGATTACAAAGCTAA 6484
Db	5859	ACTGGCCAGGCGACAGGATTCCTCTA---CTATGGGACCTCATCCTCGCATGCTCT 5915
QY	6485	TGAGAGATGTTCAATAGATGACATCAGGATCTGAAATTT-----GAAAGC 6535
Db	5916	CAGAGTCCGCTCTGCTGATGGACCATCCGGAGTTTAAGATATGTGTGTGAGGAACAGC 5975
QY	6536	ATGGGGATTGCATGCTGTTTCGACAGTGAATCTTTACTGCTGCGCCAGAGCATGGGGTTC 6595

Db	5976	TTGATGTCACACAGACTCATCCGCCACTTTCACTATACGGTGGCCAGACCACTGGAGTCC 6035
QY	6596	CTGAGAACAGCGCCCTCTAAATTCACCTTTGTGAAGTTGTTTTCGAG-----CAAGCAGGG 6649
Db	6036	CAGAAACACACCCAGTCTCTGATCCAGTTTGTGAGAACTGTCAAGGACTACATCAACAGAA 6095
QY	6650	CACATGACACACACACACTATGATTGTTTCACTGCACTGCTGGAGTGTGAAGAACTGGAGTTT 6709
Db	6096	GCCTGGGTCTGGGCCCACTGTGTGCACTGCACTGCTGTGGTAGGACTGGAACCT 6155
QY	6710	TTATTGCTCTGCAACCATTTAAACACATATAAATGACCATGATTTTGTGGATATATATG 6769
Db	6156	TTATTGCAATGGACCGAATCCCTCCAGCAGTTAGACTCCAAAGACTCTGTGGACATTTATG 6215
QY	6770	GACTAGTAGCTAACTGAGAACTGAAGAACTGATGTCAGTGCAGAACTCTGGCAGACATATA 6829
Db	6216	GAGCAGTGCACACCTTAAGACTTCACAGGTTTCATGCTCCAGACTGAGTGTCTAGTATG 6275
QY	6830	TCCTTTTACACCACTGCATCTCTGGATCTCTT 6860
Db	6276	TCTACTACATCAGTGTGTAAGAGATGCTCT 6306
BHL20957 547 bp DNA linear GSS 19-JUL-2001		
RPCI-24-230D2-TV RPCI-24 Mus musculus genomic clone RPCI-24-230D2,		
genomic survey sequence.		
BHL20957		
BHL20957.1 GI:14964469		
GSS.		
Mus musculus (house mouse)		
Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
1 (bases 1 to 547)		
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,		
Tsegaye, G., Geer, K., Krol, M., Shvartabeyn, A., Gebregeorgis, E.,		
Russell, D., de Jong, P. and Fraser, C.M.		
Mouse BAC End Sequences from Library RPCI-24		
Unpublished (1999)		
Other GSSs: RPCI-24-230D2.TJ		
Contact: Shaying Zhao		
Department of Eukaryotic Genomics		
The Institute for Genomic Research		
9712 Medical Center Dr., Rockville, MD 20850, USA		
Tel: 301 838 0200		
Fax: 301 838 0208		
Email: szhao@tigr.org		
Clones are derived from the mouse BAC library RPCI-24. For BAC		
library availability, please contact Pieter de Jong		
(pdejong@mail.cho.org). Clones may be purchased from BACPAC		
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end		
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html		
Plate: 230 row: D column: 2		
Seq primer: T7		
Class: BAC ends.		
Location/Qualifiers		
1. .547		
/organism="Mus musculus"		
/mol_type="genomic DNA"		
/strain="C57BL/6J"		
/db_xref="taxon:10090"		
/clone="RPCI-24-230D2"		
/sex="Male"		
/cell_type="Spleen/Brain"		
/clone_lib="RPCI-24"		
/note="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1;		
RPCI-24 Mouse BAC library produced by Pieter de Jong. The		
library was cloned in the pTARBAC1 cloning vector at the		
BamH1 sites using MboI partially digested male C57BL/6J		
DNA."		
FEATURES		
source		

ORIGIN

Query Match 2.5%; Score 180.6; DB 8; Length 547;
Best Local Similarity 87.2%; Pred. No. 1.8e-31;
Matches 198; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy	5493	AGCTCAGCATGATGGAAATGTAACAAAGTGGTATGTCATATTTTAATAAGCAAGGCC	5552
Db	444	AGCTCAGCAAGACGGAAATGTGACAAAGTGGTATGTCATATTTTAACAAAGCAAGGCC	385
Qy	5553	ATATTTTACAAATGAAGGCTTTCCTAACCCCTCCATGTACAGAGGAAAGCAAAAGTTTAG	5612
Db	384	ATATTTTACAAACGAAGGATTCCTTAATCCCCGTGTATAGAGGAAAGACTAAGTTCAG	325
Qy	5613	TGCCAATGAAGAAATCTACATCATAGGTGCTGTAATGTCATGTCATGCTCGCAATGA	5672
Db	324	CGGTAATGAAGAAATCTATGTGATAGGTGCTGATAATGCCCTGCATCCCTGGAATGA	265
Qy	5673	AGACAAATTTGCAATGGACCACTGAAACCAAAAAAGCAATATTAT	5719
Db	264	GGAGAAATTTGCAATGGACCTCTGAACCCCAAAAAAGCAGTATTGT	218

Search completed: July 2, 2005, 21:12:32
Job time : 32719 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:40:39 ; Search time 238 Seconds
(without alignments)
3722.975 Million cell updates/sec

Title: US-10-673-885-2
Perfect score: 11985
Sequence: 1 MDPLIFLLFLICTSTQVD.....MDSLDMEGDVLEWEETM 2291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11985	100.0	2291	6 ABP60057	Abp60057 Human pro
2	11915	99.4	2299	8 ADH41615	Adh41615 Novel hum
3	11905	99.3	2299	6 AAE37320	Aae37320 Human rec
4	11807.5	98.5	2300	5 AAO18738	Aao18738 Human NOV
5	11807.5	98.5	2300	8 ADH41625	Adh41625 Novel hum
6	11797	98.4	2299	6 AAG79724	Aag79724 Human KPP
7	11191	93.4	2281	5 AAO18736	Aao18736 Human NOV
8	11191	93.4	2281	8 ADH41619	Adh41619 Novel hum
9	11191	93.4	2281	8 ADH41613	Adh41613 Novel hum
10	10536	87.9	2301	6 ABP60058	Abp60058 Human pro
11	9874	82.4	1959	6 AAE37322	Aae37322 Human rec
12	4661	38.9	909	8 ADH41621	Adh41621 Novel hum
13	4488	37.4	855	5 AAO18737	Aao18737 Human NOV
14	4482	37.4	851	8 ADH41623	Adh41623 Novel hum
15	3804	31.7	740	8 ADH41617	Adh41617 Novel hum
16	1713.5	14.3	322	2 AAE60877	Aae60877 Product o
17	1539.5	12.8	330	6 AAE37321	Aae37321 Human rec
18	1327	11.1	2037	4 ABB71928	Abb71928 Drosophil
19	1284.5	10.7	459	8 ADP30331	Adp30331 Human sec
20	1188	9.9	1912	8 ADRA40183	Adra40183 Human pro
21	1160	9.7	1767	4 ABB62726	Abb62726 Drosophil
22	1158.5	9.7	1907	4 AAU14143	Aau14143 Human nov
23	1155.5	9.6	1897	3 AAY81785	Aay81785 Human pro
24	1155.5	9.6	1897	3 AAY56100	Aay56100 LAR tyros
25	1155.5	9.6	1897	3 AAB19712	Aab19712 Human pro

26	1155.5	9.6	1897	7 ADD18740	Add18740 Human dis
27	1155.5	9.6	1897	8 ADJ33670	Adj33670 Human leu
28	1155.5	9.6	1897	8 ADP18674	Adp18674 Human pro
29	1154	9.6	1911	2 AAR71726	Aar71726 Human PTP
30	1154	9.6	1911	2 AAW27225	Aaw27225 Human pro
31	1154	9.6	1911	2 AAW94027	Aaw94027 Human pro
32	1154	9.6	1911	2 AAU01459	Aau01459 Human pro
33	1142.5	9.5	1948	7 ADD18742	Add18742 Human dis
34	1141.5	9.5	1647	7 ABB58428	Abb58428 Drosophil
35	1141.5	9.5	1949	7 ADE57117	Ades7117 Human pro
36	1141.5	9.5	1949	7 ADE57121	Ades7121 Human pro
37	1141.5	9.5	1949	7 ADD47019	Add47019 Human pro
38	1141.5	9.5	1949	7 ADD47015	Add47015 Human pro
39	1126	9.4	1998	7 AAO24268	Aao24268 Murine va
40	1109.5	9.3	1904	5 ABB57100	Abb57100 Mouse isc
41	1109	9.3	1997	5 AAE20278	Aae20278 Human lun
42	1108	9.2	1997	7 ABR57179	Abri57179 Human PTP
43	1108	9.2	1997	7 ADJ70329	Adj70329 Human hea
44	1107	9.2	1997	3 AAB19774	Aab19774 Human pro
45	1107	9.2	1997	4 AAM78821	Aam78821 Human pro

ALIGNMENTS

RESULT 1
ABP60057
ID ABP60057 standard; protein; 2291 AA.

XX AC ABP60057;

XX DT 21-MAR-2003 (first entry)

XX DE Human phosphatase protein SEQ ID #2.

XX KW Human; phosphatase; enzyme; chromosome 12; colon adenocarcinoma;
XX KW Placenta; gene therapy.

XX OS Homo sapiens.

XX PN WO200279452-A2.

XX PD 10-OCT-2002.

XX PF 02-APR-2002; 2002WO-US009992.

XX PR 02-APR-2001; 2001US-00822871.

XX PA (PEKE) PE CORP NY.

XX PI Webster M, Wei M, Di Francesco V, Beasley EM;

XX DR WPI; 2003-046809/04.

XX DR N-PSDB; AB223703.

XX PT New human phosphatase peptides and nucleic acids encoding the peptides,
XX PT useful as models in developing human therapeutic targets, in identifying
XX PT therapeutic proteins or modulators of protein activity, and in gene
XX PT therapy.

XX PS Claim 1a; Fig 2; 220pp; English.

XX CC The invention relates to an isolated human phosphatase peptide. The human
XX CC phosphatase peptides and nucleic acids are useful as models for the
XX CC development of human therapeutic targets and in the identification of
XX CC therapeutic proteins. They also serve as targets for the development of
XX CC human therapeutic agents that modulate phosphatase activity in cells and
XX CC tissues that express the phosphatase. Experimental data indicates that
XX CC the cDNA is expressed in colon adenocarcinoma and placenta. The proteins
XX CC can be used to raise or to elicit another immune response, as a reagent
XX CC in assays designed to determine the levels of the protein in biological
XX CC fluids, as markers for tissue in which the corresponding protein is
XX CC preferentially expressed, in the identification of modulators of the

CC peptides, and in pharmacogenomic analysis. The polypeptides and
CC polynucleotides may be used in gene therapy, and as antisense
CC constructs to control phosphatase gene expression in cells, tissue and
CC organisms. The gene encoding the phosphatase of the invention is located
CC on human chromosome 12. The current sequence represents the human
CC phosphatase protein of the invention
XX
SQ Sequence 2291 AA;

Query Match 100.0%; Score 11985; DB 6; Length 2291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPLIFLLFLFTGSETQVDVSNVPGTRYDITISSITTYTSPVTRIIVNTKPGPV 60
DB 1 MDPLIFLLFLFTGSETQVDVSNVPGTRYDITISSITTYTSPVTRIIVNTKPGPV 60

QY 61 FLAGERVGSAGILLSNWTPNPNNGRIISIVIKYKEVCPMMQVYTVQVRKPDSEVLFTN 120
DB 61 FLAGERVGSAGILLSNWTPNPNNGRIISIVIKYKEVCPMMQVYTVQVRKPDSEVLFTN 120

QY 121 LNPGETTYEIKVAENSGAGVPSDPELPOTASSAPGKVNLTVYANASAVKLIWYLPKQ 180
DB 121 LNPGETTYEIKVAENSGAGVPSDPELPOTASSAPGKVNLTVYANASAVKLIWYLPKQ 180

QY 181 PNGKITSPKISVKHARSGIIVKDVSRVEDILTGKLPCECNENSESFLWSTASPSPLGRV 240
DB 181 PNGKITSPKISVKHARSGIIVKDVSRVEDILTGKLPCECNENSESFLWSTASPSPLGRV 240

QY 241 TTPSRTHSSSTLTQNEISSVWKEPISFVVTTHLRPYTYTYLFEVSAATTEAGYIDSTIVRT 300
DB 241 TTPSRTHSSSTLTQNEISSVWKEPISFVVTTHLRPYTYTYLFEVSAATTEAGYIDSTIVRT 300

QY 301 PSVPPEGPQNCVTGNITGKSPSILMDPPTIVTGKFSYRVELYGSGRILDNSTKOLKPA 360
DB 301 PSVPPEGPQNCVTGNITGKSPSILMDPPTIVTGKFSYRVELYGSGRILDNSTKOLKPA 360

QY 361 FTNLTPTMYDVYIAAETAGTGPKSNISVFPPDVPDGAVIDLQLAESTVQRIITWKKP 420
DB 361 FTNLTPTMYDVYIAAETAGTGPKSNISVFPPDVPDGAVIDLQLAESTVQRIITWKKP 420

QY 421 RQPNGLIINQVRKVLVPETGIIENLTLLGNNEYINDMPAPEIIVNVEPMVGLYEGSAEM 480
DB 421 RQPNGLIINQVRKVLVPETGIIENLTLLGNNEYINDMPAPEIIVNVEPMVGLYEGSAEM 480

QY 481 SSDLHSLATFIYNHSPDKNPPARNRAEDQTSVPVTRNQVITDIAEQLSYVIRLVPFT 540
DB 481 SSDLHSLATFIYNHSPDKNPPARNRAEDQTSVPVTRNQVITDIAEQLSYVIRLVPFT 540

QY 541 EHMISVSAFTIMEGEGPTVLSVTRQOVPSIKIINYKNISSSILLYWDPPPEYNGKIT 600
DB 541 EHMISVSAFTIMEGEGPTVLSVTRQOVPSIKIINYKNISSSILLYWDPPPEYNGKIT 600

QY 601 HYTIYAMELDTNRAFOITIDNSFLITGLKKYTKYKWRVAASTHDESSISENDIFVRT 660
DB 601 HYTIYAMELDTNRAFOITIDNSFLITGLKKYTKYKWRVAASTHDESSISENDIFVRT 660

QY 661 SEDEPESSQDVEIVDVTADEIRLKWSPPEKPNGLIIAYEVLVKNITDLYMKNSTTDDII 720
DB 661 SEDEPESSQDVEIVDVTADEIRLKWSPPEKPNGLIIAYEVLVKNITDLYMKNSTTDDII 720

QY 721 LRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGELELSP 780
DB 721 LRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGELELSP 780

QY 781 LPSSPNGLIIKKYTYLKRSGNNEERTINTTSLTQNIKVLKKYQVYIIEVSASTLKGEGV 840
DB 781 LPSSPNGLIIKKYTYLKRSGNNEERTINTTSLTQNIKVLKKYQVYIIEVSASTLKGEGV 840

QY 841 RSAPISILTEEDAPSPQDPSVKQLSGVTVKLSWQPPLEPNGLIILYYTVVWNRSSLKT 900
DB 841 RSAPISILTEEDAPSPQDPSVKQLSGVTVKLSWQPPLEPNGLIILYYTVVWNRSSLKT 900

QY 901 INVTETSLSDLDYNNVBSAYVTASTRFGDKTGSNIISFQTPEGASDPDPKDYVYANL 960
DB 901 INVTETSLSDLDYNNVBSAYVTASTRFGDKTGSNIISFQTPEGASDPDPKDYVYANL 960

QY 961 SSSSIILFWTPPSKNGIIQYYSVYVRYNTSGTFMQNFTLHETLNDPDMVTSTIIDKLT 1020
DB 961 SSSSIILFWTPPSKNGIIQYYSVYVRYNTSGTFMQNFTLHETLNDPDMVTSTIIDKLT 1020

QY 1021 PSYTFMLTASTSVGNKSSDIIEVYTDODIPEGFVGNLTYESISSSTAINVSWPPAPQ 1080
DB 1021 PSYTFMLTASTSVGNKSSDIIEVYTDODIPEGFVGNLTYESISSSTAINVSWPPAPQ 1080

QY 1081 NGLVYVYVSLILQOQPRHVRPLVYERSIYFDNLEKTYDYLKILTPSTKEGFSDTYTAQ 1140
DB 1081 NGLVYVYVSLILQOQPRHVRPLVYERSIYFDNLEKTYDYLKILTPSTKEGFSDTYTAQ 1140

QY 1141 LYIKTEEDVPETSPIINTPKNLSSTVLLSWDPPVKPNGAIIISYDLTLQGPNEYSFITS 1200
DB 1141 LYIKTEEDVPETSPIINTPKNLSSTVLLSWDPPVKPNGAIIISYDLTLQGPNEYSFITS 1200

QY 1201 DNYIILELSPEFLYSFAAARTRKGLGPSSILFEYTDSEVPLAPPONLTILINCTSDFW 1260
DB 1201 DNYIILELSPEFLYSFAAARTRKGLGPSSILFEYTDSEVPLAPPONLTILINCTSDFW 1260

QY 1261 LKWSPSPLPGGIVKYVSPKIHHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIRVSAFTK 1320
DB 1261 LKWSPSPLPGGIVKYVSPKIHHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIRVSAFTK 1320

QY 1321 VGNQGFQSNVVKFTQESVDPVQVNMCMATSWQSVLWKWDPPKANGIIITOYMTVVERN 1380
DB 1321 VGNQGFQSNVVKFTQESVDPVQVNMCMATSWQSVLWKWDPPKANGIIITOYMTVVERN 1380

QY 1381 STKVSPODHMYTFIILLANTSVYKVRASTSAGEDESTCHVSTLPETVPSVPTNIAFSD 1440
DB 1381 STKVSPODHMYTFIILLANTSVYKVRASTSAGEDESTCHVSTLPETVPSVPTNIAFSD 1440

QY 1441 VQSTSATLTWIRPDITLGYFQNYKITTLQRAQCKEWESEECVEYQKIQYLYEAHLTEFT 1500
DB 1441 VQSTSATLTWIRPDITLGYFQNYKITTLQRAQCKEWESEECVEYQKIQYLYEAHLTEFT 1500

QY 1501 VTGLKKFWRYPQVAASNAGYGNASNMISTKTLPGPDGPPENHVHVSATSPSISWS 1560
DB 1501 VTGLKKFWRYPQVAASNAGYGNASNMISTKTLPGPDGPPENHVHVSATSPSISWS 1560

QY 1561 EBAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIETIKOLEIFTRXSVLITAFNGNIS 1620
DB 1561 EBAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIETIKOLEIFTRXSVLITAFNGNIS 1620

QY 1621 AAYVEGKSAEMI VTTLESAPKDPNNMTFQKI PDEVTKFQTLFPLPPSQPNNGNIQVYQAL 1680
DB 1621 AAYVEGKSAEMI VTTLESAPKDPNNMTFQKI PDEVTKFQTLFPLPPSQPNNGNIQVYQAL 1680

QY 1681 VYREDDPTAVQIHNSIIQKNTTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPMRITM 1740
DB 1681 VYREDDPTAVQIHNSIIQKNTTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPMRITM 1740

QY 1741 DIKAPARKTKPTPIYDATGKLLVSTTTITRMPICYSDDDHGPKNQVOLATEGAQHD 1800
DB 1741 DIKAPARKTKPTPIYDATGKLLVSTTTITRMPICYSDDDHGPKNQVOLATEGAQHD 1800

QY 1801 GNVTKWYDAYFNKARPYFTNEGPNPPCTEGTKFSGNEEIIYIIGADNACMIPGNEDKTC 1860
DB 1801 GNVTKWYDAYFNKARPYFTNEGPNPPCTEGTKFSGNEEIIYIIGADNACMIPGNEDKTC 1860

QY 1861 NGPLKPKQYLPKFRATNIMGQFTDSYSDPDKTGBGLSERTVEILSVTCILSIIILL 1920
DB 1861 NGPLKPKQYLPKFRATNIMGQFTDSYSDPDKTGBGLSERTVEILSVTCILSIIILL 1920

QY 1921 GTAI FAFARIKOKEGGTYSPQDAEIIDTKLQDOLITVADLEKDEBLTRPISKKSFL 1980
DB 1921 GTAI FAFARIKOKEGGTYSPQDAEIIDTKLQDOLITVADLEKDEBLTRPISKKSFL 1980

QY 1981 QHVEELCTNNLKFQEEFSELPKFLQDLSSTDADLPWNRKRNFPNIKFPYNNNNRVKLI 2040

Db 1981 QHVEELCTNNLKFQSEFSELPKFLQDLSSTDADLPWNRKAKRFPNKNNNNNRKLIA 2040
QY 2041 DASVPGSDYINASYISYLCYCPNEFIATQGPLCTGVDFWRWVWETRAKTLVMLTQCFEKG 2100
Db 2041 DASVPGSDYINASYISYLCYCPNEFIATQGPLCTGVDFWRWVWETRAKTLVMLTQCFEKG 2100
QY 2101 RIRCHOYWPEDNKPVTVPFGDIVITKLMEDVQIDMTIRDLKIERHGDGCMVTRQCNFTAWPE 2160
Db 2101 RIRCHOYWPEDNKPVTVPFGDIVITKLMEDVQIDMTIRDLKIERHGDGCMVTRQCNFTAWPE 2160
QY 2161 HGVPENSAPLHFVKLVARSRAHDTTPMLVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220
Db 2161 HGVPENSAPLHFVKLVARSRAHDTTPMLVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220
QY 2221 YGLVABLRGSRMCMVQLAQYIFLHQCILDLSNKGSNQPICFVNYSAIQKMSDLSIDAMEG 2280
Db 2221 YGLVABLRGSRMCMVQLAQYIFLHQCILDLSNKGSNQPICFVNYSAIQKMSDLSIDAMEG 2280
QY 2281 DVELEWBEETM 2291
Db 2281 DVELEWBEETM 2291
RESULT 2
ADH41615
ID ADH41615 standard; protein; 2299 AA.
AC ADH41615;
XX
DT 25-MAR-2004 (first entry)
DE Novel human protein NOV15b.
XX
KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
KW antiparkinsonian; antiaschmatic; antiinfertility; cardiomyopathy;
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX
OS Homo sapiens.
XX
PN WO2003102159-A2.
XX
PD 11-DEC-2003.
XX
PF 04-JUN-2003; 2003WO-US017573.
XX
PR 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386864P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387866P.
PR 11-JUN-2002; 2002US-0387806P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.

PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403450P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;
PI Ellerman K, Rittenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;
XX
XX WPI; 2004-053467/05.
DR N-PSDB; ADH41614.
XX
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.
XX
PS Claim 2; SEQ ID NO 168; 1503pp; English.
XX
XX The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
XX example of the polypeptide of the invention.
SQ Sequence 2299 AA;

Query Match		99.4%;	Score 11915;	DB 8;	Length 2299;	
Best Local Similarity		99.2%;	Pred. No. 0;			
Matches 2282;		Conservative	3;	Mismatches	5;	Indels 10; Gaps 2;
QY	1	MDPLIIFLLFGTSETQVDNSVNPVGTGYDITISSITTYTSPVTRIVTPNVTKPGPPV	60			
DB	1	MDPLIIFLLFGTSETQVDNSVNPVGTGYDITISSITTYTSPVTRIVTPNVTKPGPPV	60			
QY	61	FLAGERVGAGILLSNWTPNPNGRIISIVIKYKEVCPMWTQYTVQVRKPSDLEVLN	120			
DB	61	FLAGERVGAGILLSNWTPNPNGRIISIVIKYKEVCPMWTQYTVQVRKPSDLEVLN	120			
QY	121	LNPGTTYEIKVAENAGIGVSDPELFOFASAPGVNLTVEANASAVKLIWLPQ	180			
DB	121	LNPGTTYEIKVAENAGIGVSDPELFOFASAPGVNLTVEANASAVKLIWLPQ	180			
QY	181	PNKITSFKISVKHARSIGVVDVSRVEDILTGLKPECNENSESFLWSTASPSPTLGRV	240			
DB	181	PNKITSFKISVKHARSIGVVDVSRVEDILTGLKPECNENSESFLWSTASPSPTLGRV	240			
QY	241	TPPSRTHSSSLTQNEISSVWKEPISFVTHLRPTTYTLFEVSAATTEAGYIDSTIVRT	300			
DB	241	TPPSRTHSSSLTQNEISSVWKEPISFVTHLRPTTYTLFEVSAATTEAGYIDSTIVRT	300			
QY	301	PESVPEGPPQNCVTGNI TGKFSILWDPTITVTKFESYRVELYGPGRILDNSTDKLKEA	360			
DB	301	PESVPEGPPQNCVTGNI TGKFSILWDPTITVTKFESYRVELYGPGRILDNSTDKLKEA	360			
QY	361	FTNLTPFTMYDVYIAAETSAGTGPKSNI SVFTPPDVPVGA VFDQLAEVESTQVRI TWKCP	420			
DB	361	FTNLTPFTMYDVYIAAETSAGTGPKSNI SVFTPPDVPVGA VFDQLAEVESTQVRI TWKCP	420			
QY	421	RQNGIINOYRVKLVPEPTGII LENTLLTGNEIYINDPMAPEI VNI VEPMVGLYEGSAEM	480			
DB	421	RQNGIINOYRVKLVPEPTGII LENTLLTGNEIYINDPMAPEI VNI VEPMVGLYEGSAEM	480			
QY	481	SSDLHSLATFIYNSHPDKNPPARNRAEDQTS PVVTRNOYITDIAEQLSYVIRLVPET	540			
DB	481	SSDLHSLATFIYNSHPDKNPPARNRAEDQTS PVVTRNOYITDIAEQLSYVIRLVPET	540			
QY	541	EHMISVSAFTIMEGEGPTVLVSRTRQVPS SIKI INYKNI SSSSILLYNDPPEYNGKIT	600			
DB	541	EHMISVSAFTIMEGEGPTVLVSRTRQVPS SIKI INYKNI SSSSILLYNDPPEYNGKIT	600			
QY	601	HYTIYAMELDTNRAFOITIDNSFLITGLKKTYYKORVAASTHDGESSLSBENDIFVRT	660			
DB	601	HYTIYAMELDTNRAFOITIDNSFLITGLKKTYYKORVAASTHDGESSLSBENDIFVRT	660			
QY	661	SEDEPESSQDVEVIDVDTADEIRL KWS PPEKNGI IIAEVLKYKNIDTLYMKNTSTTDII	720			
DB	661	SEDEPESSQDVEVIDVDTADEIRL KWS PPEKNGI IIAEVLKYKNIDTLYMKNTSTTDII	720			
QY	721	LRNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNI SSGEIELSF	780			
DB	721	LRNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNI SSGEIELSF	780			
QY	781	LPPSSPNGII KKYTIYLRKSGNEERTINTTSQNI KVL KKYQYII EVSASTLKGEV	840			
DB	781	LPPSSPNGII KKYTIYLRKSGNEERTINTTSQNI KVL KKYQYII EVSASTLKGEV	840			
QY	841	RSAPISILTEEDAPSDPPQDPSVKQLSGVTYKLSWQPPLEPGCI IILYTYVYVWNRSSSLKT	900			
DB	841	RSAPISILTEEDAPSDPPQDPSVKQLSGVTYKLSWQPPLEPGCI IILYTYVYVWNRSSSLKT	900			
QY	901	INVTETSLSDLDYNVEYSAYTASTRFGDGKTSNII SFQTPPEGAPSDPPKDVYANL	960			
DB	901	INVTETSLSDLDYNVEYSAYTASTRFGDGKTSNII SFQTPPEGAPSDPPKDVYANL	960			
QY	961	SSSSIIILFWTPPSKPNGLII QYYSVYVRNTSGTFMQNFTHLNTDNDNNMTVSTIIDKLTJ	1020			
DB	961	SSSSIIILFWTPPSKPNGLII QYYSVYVRNTSGTFMQNFTHLNTDNDNNMTVSTIIDKLTJ	1020			
QY	1021	FSYTYFWLTASTSVGNKSSDII IEVYTDQDIPGFGVGNLTYESISSSTAINVSWPPAQP	1080			
DB	1021	FSYTYFWLTASTSVGNKSSDII IEVYTDQDIPGFGVGNLTYESISSSTAINVSWPPAQP	1080			
QY	1081	NGLVFYVSLILQOQPRHVRPPLVYERSIYPDNI LEKYTDYILKITPSTKGFSDTYTAQ	1140			
DB	1081	NGLVFYVSLILQOQPRHVRPPLVYERSIYPDNI LEKYTDYILKITPSTKGFSDTYTAQ	1140			
QY	1141	LYIKTEEDVPETSPIINTFKNLSSVLLSWDPVVKPKNAGII SYDLTLQGNENYSFITS	1200			
DB	1141	LYIKTEEDVPETSPIINTFKNLSSVLLSWDPVVKPKNAGII SYDLTLQGNENYSFITS	1200			
QY	1201	DNYIILBELSPPTLYSFFAAARTRKGLGPSSILFFYTDESVP LAPPQNLTLINCTSDFW	1260			
DB	1201	DNYIILBELSPPTLYSFFAAARTRKGLGPSSILFFYTDESVP LAPPQNLTLINCTSDFW	1260			
QY	1261	LKWSPSPLPGGIVKYVSFKIHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIRVSAFTK	1320			
DB	1261	LKWSPSPLPGGIVKYVSFKIHEHETDTIYKNI SGFKTEAKLVGLEPVSTYSIRVSAFTK	1320			
QY	1321	VGNQGFNSNVKFTTQESVPDVVNQMCWATSWQSVLVKWDPPKKANGII TOYMYTVERN	1380			
DB	1321	VGNQGFNSNVKFTTQESVPDVVNQMCWATSWQSVLVKWDPPKKANGII TOYMYTVERN	1380			
QY	1381	STKVSPODHMYTIFIKLLANTS YVFKVRASTASAGEDESTCHVSTLPETVPSVPTNIAFSD	1440			
DB	1381	STKVSPODHMYTIFIKLLANTS YVFKVRASTASAGEDESTCHVSTLPETVPSVPTNIAFSD	1440			
QY	1441	VOSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESEECVEYQKI QYLYEAHLTBEET	1500			
DB	1441	VOSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESEECVEYQKI QYLYEAHLTBEET	1500			
QY	1501	YVGLKPFWRYPQVAASTNAGVGNASNISYKTLPGPDGPPENVHVATSPFSISWS	1560			
DB	1501	YVGLKPFWRYPQVAASTNAGVGNASNISYKTLPGPDGPPENVHVATSPFSISWS	1560			
QY	1561	BEAVITGPTCYLIDVKSVDNDEFNISFKISNEENKTI EIKOLEIFTRYSVVITATGNIS	1620			
DB	1561	BEAVITGPTCYLIDVKSVDNDEFNISFKISNEENKTI EIKOLEIFTRYSVVITATGNIS	1620			
QY	1621	AAVTEGKSAAEMIVTTLESAPKDPNNMTFKI PDEVTKFQLTFLPPSOPNGNI OYVQAL	1680			
DB	1621	AAVTEGKSAAEMIVTTLESAPKDPNNMTFKI PDEVTKFQLTFLPPSOPNGNI OYVQAL	1680			
QY	1681	VYREDDPTAVQIHNLSII OKTNTFVIAMLEGLKGHTYNI SVYVNSAGAGKVPMTIM	1740			
DB	1681	VYREDDPTAVQIHNLSII OKTNTFVIAMLEGLKGHTYNI SVYVNSAGAGKVPMTIM	1740			
QY	1741	DIKAPARPKTKPTPIYDATGKLLVTSTTTIRMPICYSDDDHGP IKNVQLATETGAQHD	1800			
DB	1741	DIKAPARPKTKPTPIYDATGKLLVTSTTTIRMPICYSDDDHGP IKNVQLATETGAQHD	1800			
QY	1801	GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYII GADNACMI PGNEDKTC	1860			
DB	1801	GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYII GADNACMI PGNEDKTC	1860			
QY	1861	NGPLKPKQYLPKPFATNIMGOTDSYSDPVKTLGEGLSERTVEI ILSVTLICILSILL	1920			
DB	1861	NGPLKPKQYLPKPFATNIMGOTDSYSDPVKTLGEGLSERTVEI ILSVTLICILSILL	1920			
QY	1921	GTAIFAFARIROKQEGGTYSPQDAEIIDTKLKDQLITVADLELKDRLTR - - - - -	1972			
DB	1921	GTAIFAFARIROKQEGGTYSPQDAEIIDTKLKDQLITVADLELKDRLTR - - - - -	1972			
QY	1973	-PISKKSFLQHVVEELCTNNNLKFQBEFSEL PKFLODLSSTDADLPWNAKNRFPNIKPN	2031			
DB	1981	KPI SKKSFLQHVVEELCTNNNLKFQBEFSEL PKFLODLSSTDADLPWNAKNRFPNIKPY -	2039			
QY	2032	NNNRVKLTADASVPGSDYINASYISGYLCPNEFIATQGPLCTGDFWRVWVETRAKTIV	2091			
DB	2040	NNNRVKLTADASVPGSDYINASYISGYLCPNEFIATQGPLCTGDFWRVWVETRAKTIV	2099			
QY	2092	MLTQCFEKGRIRCHOYWPEDNKPVTVFGDIVITIKLMEDVQIDWTIRDLKIERHGDQMTVR	2151			

Db 2100 MLTQCFEKGIRCHQWPENKPVTVFGDIVITVKLMEDVQIDMTIRDLKIERHGDCTVR 2159
Qy 2152 QCNFTAWPEHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGTFVIALDHLTOH 2211
Db 2160 QCNFTAWPEHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGTFVIALDHLTOH 2219
Qy 2212 INDHDFVDIYGLVAELSRMCMQVNLQAYIFLHQICILDLNKGNSQNPICFVNSALOK 2271
Db 2220 INDHDFVDIYGLVAELSRMCMQVNLQAYIFLHQICILDLNKGNSQNPICFVNSALOK 2279
Qy 2272 MDSLDAWEGDVELEWBEETM 2291
Db 2280 MDSLDAWEGDVELEWBEETM 2299

RESULT 3

AAE37320
ID AAE37320 standard; protein; 2299 AA.

XX AAE37320;

XX 07-AUG-2003 (first entry)

XX Human receptor tyrosine kinase protein #1.

XX Human; receptor tyrosine phosphatase; diabetes; obesity; CNS disorder;
KW multiple sclerosis; epilepsy; chronic obstructive pulmonary disease;
KW COPD; Parkinson's disease; stroke; cardiovascular disorder; arrhythmia;
KW congestive heart failure; myocardial infarction; ischaemic heart disease;
KW gene therapy; anorectic; cardiact; neuroprotective; anticonvulsant;
KW cerebroprotective; vasotropic; antiarrhythmic; receptor; enzyme;
KW chromosome 12.

XX Homo sapiens.

XX WO2003033688-A1.

PN 24-APR-2003.

PD 14-OCT-2002; 2002WO-EP011473.

PP 16-OCT-2001; 2001US-0329329P.

XX (PARB) BAYER AG.

XX Liou J;

XX WPI; 2003-403215/38.

DR N-PSDB; AAD56411.

XX Novel polynucleotides encoding human receptor tyrosine phosphatase

XX polypeptides, useful for treating diabetes, CNS disorders, obesity,

XX chronic obstructive pulmonary disease and cardiovascular disorders.

XX Claim 1; Fig 2; 163pp; English.

XX The present invention relates to receptor tyrosine phosphatase proteins
CC and polynucleotides encoding them. Sequences of the invention are useful
CC in the preparation of medicaments for modulating the activity of receptor
CC tyrosine phosphatase in disease such as diabetes, obesity, CNS disorders
CC (multiple sclerosis, epilepsy, Parkinson's disease and stroke), chronic
CC obstructive pulmonary disease (COPD) and cardiovascular disorders (e.g.
CC congestive heart failure, myocardial infarction, ischaemic heart disease
CC and arrhythmia). They are also used in gene therapy. The present sequence
CC is human receptor tyrosine kinase protein. Receptor tyrosine kinase gene
CC is located on chromosome 12

XX Sequence 2299 AA;

XX Query Match 99.3%; Score 11905; DB 6; Length 2299;

XX Best Local Similarity 99.2%; Pred. No. 0;

XX Matches 2282; Conservative 3; Mismatches 5; Indels 10; Gaps 2;

Qy 1 MDPLIFLLFTGTSETQVDSNVVPGTRYDITISSISTYTSPTVTRIVTPNVTXPGPV 60
Db 1 MDPLIFLLFTGTSETQVDSNVVPGTRYDITISSISTYTSPTVTRIVTPNVTXPGPV 60
Qy 61 FLAGERVGSAGILLSNWTPPNPGRISIVIVKYKEVCPMMQVTVTVRSKPSDSLEVLN 120
Db 61 FLAGERVGSAGILLSNWTPPNPGRISIVIVKYKEVCPMMQVTVTVRSKPSDSLEVLN 120
Qy 121 LNPFGTYEIKVAEENSAGIGVFSDDPLFQTABSAFGKVVNLTVAYNANASAVKLIWYLP 180
Db 121 LNPFGTYEIKVAEENSAGIGVFSDDPLFQTABSAFGKVVNLTVAYNANASAVKLIWYLP 180
Qy 181 PNGKITSPKISVKHARSIGVIVKDVSIKVEDILTKLPESNENSESFLWSTASPSPTLGRV 240
Db 181 PNGKITSPKISVKHARSIGVIVKDVSIKVEDILTKLPESNENSESFLWSTASPSPTLGRV 240
Qy 241 TPPSRTTHSSSTLTQNEISSVWKEPISFVVTTHLRPTTYLFEVSAAATTEAGVIDSTIVRT 300
Db 241 TPPSRTTHSSSTLTQNEISSVWKEPISFVVTTHLRPTTYLFEVSAAATTEAGVIDSTIVRT 300
Qy 301 PESVPEGPQNCVTGNTGKSPSILWDPPTIVTGKFSYRVELYGPGRILDNSTDKLKA 360
Db 301 PESVPEGPQNCVTGNTGKSPSILWDPPTIVTGKFSYRVELYGPGRILDNSTDKLKA 360
Qy 361 FTNLTPFTMYDYIAAETSAGTGPKNISVFTPPDVGAVFDLQLAESTQVRIWTKKP 420
Db 361 FTNLTPFTMYDYIAAETSAGTGPKNISVFTPPDVGAVFDLQLAESTQVRIWTKKP 420
Qy 421 ROPNGIINQYRVKVLVPETGILENTLTGNEYINDMPAIEVINIVEMVGLYEGSASM 480
Db 421 ROPNGIINQYRVKVLVPETGILENTLTGNEYINDMPAIEVINIVEMVGLYEGSASM 480
Qy 481 SSDLSLATFIYNHSHDPKNEPARNAEQTSVVTTRNOYITDIAAEOLSYVIRLVPFT 540
Db 481 SSDLSLATFIYNHSHDPKNEPARNAEQTSVVTTRNOYITDIAAEOLSYVIRLVPFT 540
Qy 541 EHMISVSAFTIMGEGPPTVLSVTRTQQVPSSIKIINYKNISSSILLYWDPPPEYNGKIT 600
Db 541 EHMISVSAFTIMGEGPPTVLSVTRTQQVPSSIKIINYKNISSSILLYWDPPPEYNGKIT 600
Qy 601 HYTIYAMELDTNRAFOITIDNSFLITGLKKTYYKMRVAASTHOGESSLENDIFVRT 660
Db 601 HYTIYAMELDTNRAFOITIDNSFLITGLKKTYYKMRVAASTHOGESSLENDIFVRT 660
Qy 661 SEDBPESQDVEVDVDADEIRLKWSPPEKNGIIIAEVLKYKNDITLYMKNSTTIDII 720
Db 661 SEDBPESQDVEVDVDADEIRLKWSPPEKNGIIIAEVLKYKNDITLYMKNSTTIDII 720
Qy 721 LRNLPHLTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGIELSF 780
Db 721 LRNLPHLTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGIELSF 780
Qy 781 LPPSSPNGIIKKYTIYLRKSGNEERTINTSLTONIKVLKYYQIYIEVSASTYKGEV 840
Db 781 LPPSSPNGIIKKYTIYLRKSGNEERTINTSLTONIKVLKYYQIYIEVSASTYKGEV 840
Qy 841 RSAPISIIITEDADPSPPQDPSVKOLSGVTVKLSWQPLENGIILYTVVYVWNSSLKT 900
Db 841 RSAPISIIITEDADPSPPQDPSVKOLSGVTVKLSWQPLENGIILYTVVYVWNSSLKT 900
Qy 901 INVETSLESLDLYNVEYSAYVTAFTFGDKTGSNIISFOTPEGADSPDPKDYVYANL 960
Db 901 INVETSLESLDLYNVEYSAYVTAFTFGDKTGSNIISFOTPEGADSPDPKDYVYANL 960
Qy 961 SSSSIIILFWTPPSKPNGIIOYYSVYRNTSGTFMQNFTLHETLNDPDMNTVSTIIDKLT 1020
Db 961 SSSSIIILFWTPPSKPNGIIOYYSVYRNTSGTFMQNFTLHETLNDPDMNTVSTIIDKLT 1020
Qy 1021 PSYTYFWLTASTSVGNKNSDIIIEVYTDQDIPGFCVGLTYESISSSTAINVSWPPAQP 1080
Db 1021 PSYTYFWLTASTSVGNKNSDIIIEVYTDQDIPGFCVGLTYESISSSTAINVSWPPAQP 1080

1081 NGLVFYVSLILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSDYTTAQ 1140
1081 NGLVFYVSLILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSDYTTAQ 1140
1141 LYIKTEBVPETSPINTPKNSSTSVLISWPPVVPNGAIIISYDLTLQGNPNVTSFITS 1200
1141 LYIKTEBVPETSPINTPKNSSTSVLISWPPVVPNGAIIISYDLTLQGNPNVTSFITS 1200
1201 DNYIIIEELSPFTLYSFFFAAARKGLGPSSILFFYTDSEVPLAPPQNLTLINCTSDVFM 1260
1201 DNYIIIEELSPFTLYSFFFAAARKGLGPSSILFFYTDSEVPLAPPQNLTLINCTSDVFM 1260
1261 LKWSPSPLPGGIVKVSFKIHEHETDIYYKNIISGFKTEAKLVGLEPVSITYSIRVSAFTK 1320
1261 LKWSPSPLPGGIVKVSFKIHEHETDIYYKNIISGFKTEAKLVGLEPVSITYSIRVSAFTK 1320
1321 VGNQOFSNVKFTTQESVPDVVQNMCMATSWQSVLVKWDPPKANGIIITQYMTVERN 1380
1321 VGNQOFSNVKFTTQESVPDVVQNMCMATSWQSVLVKWDPPKANGIIITQYMTVERN 1380
1381 STKVSQDHYMTFIKLLANTSYVKVRASTASAGDEGSTCHVSTLPETVPSVPTNIAFSD 1440
1381 STKVSQDHYMTFIKLLANTSYVKVRASTASAGDEGSTCHVSTLPETVPSVPTNIAFSD 1440
1441 VOSTSATLWIRPDTILGYFQNYKIITQLRAQCKEWESEECVEYQKIQYLYEAHLTEST 1500
1441 VOSTSATLWIRPDTILGYFQNYKIITQLRAQCKEWESEECVEYQKIQYLYEAHLTEST 1500
1501 VYGLKKFRWYRFOVAASSTNAGYGNASNWIISTKTLPGPPDGPENVHVATSPFSISWS 1560
1501 VYGLKKFRWYRFOVAASSTNAGYGNASNWIISTKTLPGPPDGPENVHVATSPFSISWS 1560
1561 EPAVITGPTCYLIDVKSVDNDENISFIKSNEENKTIEIKOLEIFTRYSVVITVTAFTGNIS 1620
1561 EPAVITGPTCYLIDVKSVDNDENISFIKSNEENKTIEIKOLEIFTRYSVVITVTAFTGNIS 1620
1621 AAYVEGKSSAEMIVTTLESAPKDPNNMTFKQIPDEVTKFQLTFLPPSPNGNIQYQAL 1680
1621 AAYVEGKSSAEMIVTTLESAPKDPNNMTFKQIPDEVTKFQLTFLPPSPNGNIQYQAL 1680
1681 VYREDDPTAVQIHNLSIIOKTNTFVIAMLEGLKGHTYINSVYVNSAGAGKPKVPMRITM 1740
1681 VYREDDPTAVQIHNLSIIOKTNTFVIAMLEGLKGHTYINSVYVNSAGAGKPKVPMRITM 1740
1741 DIKAPARPKTPPIYDATGKLVSTTTITRMPICYSDDHGPIKNVQVLATETGAQHD 1800
1741 DIKAPARPKTPPIYDATGKLVSTTTITRMPICYSDDHGPIKNVQVLATETGAQHD 1800
1801 GNVTKWYDAYFNKARPYFTNEGPPNPPCTEGKTFSGNEEIIYIIGADNACMIPGNEDKIC 1860
1801 GNVTKWYDAYFNKARPYFTNEGPPNPPCTEGKTFSGNEEIIYIIGADNACMIPGNEDKIC 1860
1861 NGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCISILL 1920
1861 NGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCISILL 1920
1921 GTAIFAFARIQKQEGGYTSPQDAEIIITKLKLDQIITVADLELKDRLTR----- 1972
1921 GTAIFAFARIQKQEGGYTSPQDAEIIITKLKLDQIITVADLELKDRLTR----- 1972
1973 -PTSKSFLQHVBEELCTNNLKFQEBFSLPKFPLQDLSSTDADLPNWRKNRFPNPKYN 2031
1981 KPISKKSFLQHVBEELCTNNLKFQEBFSLPKFPLQDLSSTDADLPNWRKNRFPNPKY- 2039
2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWVWVETRAKTIV 2091
2040 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWVWVETRAKTIV 2099
2092 MLTQCCEKGRIRCHQWYPEDNKVTVFGDIVITIKLMEDVQIDWITRDLKIERHGDQMYR 2151
2100 MLTQCCEKGRIRCHQWYPEDNKVTVFGDIVITIKLMEDVQIDWITRDLKIERHGDQMYR 2159
2152 QCNFTAWPEHGVSPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTQH 2211

Db 2160 QCNFTAWPEHGVSPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTQH 2219
QY 2212 INDHDFVDIYGLVAELSRERMCMQNLAQYIFLHQICILDLSNKGSNOPICFVNSALQK 2271
Db 2220 INDHDFVDIYGLVAELSRERMCMQNLAQYIFLHQICILDLSNKGSNOPICFVNSALQK 2279
QY 2272 MDSLDAMEGDVELEWSETTM 2291
Db 2280 MDSLDAMEGDVELEWSETTM 2299
RESULT 4
AAO18738
ID AAO18738 standard; protein; 2300 AA.
XX AAO18738;
AC AAO18738;
XX
DT 24-OCT-2002 (first entry)
XX Human NOV2c protein.
DE Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease;
KW storage disorder; muscle disorder; neurodegenerative disorder; nootropic;
KW developmental defect; neuroprotective; antiparkinsonian; hypotensive;
KW hypertensive; haemostatic; cardiant; antitanginal; dermatological;
KW immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV;
KW antiparasitic; antiallergic; antiasthmatic; antirheumatic; antiarthritis;
KW vulnery; anorectic; antidiabetic; immunomodulator; antipsoriatic;
KW nephrotropic; kerolytic; antiulcer; cerebroprotective; anticonvulsant;
KW antifertility; antimanic; antidepressant; metabolic; cytostatic;
KW tranquilizer; analgesic.
XX Homo sapiens.
OS
XX WO200257450-A2.
XX 25-JUL-2002.
XX 29-NOV-2001; 2001WO-US048922.
XX 29-NOV-2000; 2000US-0253834P.
PR 30-NOV-2000; 2000US-0250926P.
PR 25-JAN-2001; 2001US-0264180P.
PR 20-AUG-2001; 2001US-0313656P.
PR 05-OCT-2001; 2001US-0327456P.
PR 28-NOV-2001; 2001US-00327456.
XX (CURA-) CURAGEN CORP.
PA Edinger S, Macdougall JR, Millet I, Ellerman K, Stone DJ;
PI Gerlach V, Grosse WM, Alsobrook JP, Lepley DM, Rieger D, Borgeas CE;
PI Casman SJ, Spytek KA, Boldog FL, Li L, Padigar M, Mishra V;
PI Patturajan M, Shenoy S, Rastelli L, Tchernev VT, Vernet CAM;
PI Zerhusen BD, Malyankar UM, Guo X, Miller CE, Gangolli EA;
XX WPI: 2002-590741/63.
DR N-PSDB; ABO6283.
XX
PT Novel isolated polypeptide, designated NOVX, useful for treating or
PT preventing in NOVX-associated disorders e.g. cardiomyopathy,
PT atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.
XX
PS Claim 1; Page 30; 353pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC several novel human proteins, designated NOVX. These can be used in the
CC treatment of, amongst others, cancers, autoimmune diseases, infections,
CC inflammatory diseases, storage disorders, muscle disorders,
CC neurodegenerative diseases and developmental defects. The present
CC sequence is a protein of the invention
XX
SQ Sequence 2300 AA;

Query Match	98.5%; Score 11807.5; DB 5; Length 2300;	
Best local Similarity	98.7%; Pred. No. 0;	
Matches 2276; Conservative .5; Mismatches 5; Indels 19; Gaps 9;		
QY	1 MDLFIPLFLFIGTSETQVDVSNVPGTRDYDITISSITSTYTSPTVRITVNTKPGPPV 60	1016 DKLTIFSYTFFWLTASTSVGNGKSSDIIEVVTDDIPEGFVGNLTYESISSTAINVSW 1075
DB	1 MDLFIPLFLFIGTSETQVDVSNVPGTRDYDITISSITSTYTSPTVRITVNTKPGPPV 60	1017 DKLTIFSYTFFWLTASTSVGNGKSSDIIEVVTDDIPEGFVGNLTYESISSTAINVSW 1076
QY	61 FLAGERVGSAGILLSNWTPNPNGRIISYIVKYKEVCPMMQTYTVQVRKSPDSLEVLLTN 120	1076 PPAQPNGLVYVYVSLILOQTPRHVRPPLVYIERSYFDNLEKVTYDILKITPSTEGKFS 1135
DB	61 FLAGERVGSAGILLSNWTPNPNGRIISYIVKYKEVCPMMQTYTVQVRKSPDSLEVLLTN 120	1077 PPAQPNGLVYVYVSLILOQTPRHVRPPLVYIERSYFDNLEKVTYDILKITPSTEGKFS 1136
QY	121 LNPCTTYEIKVAENAGAGVFDPLFQTAES-APGKVNLTVAYNASAVKLIWYLR 179	1136 TYTAQLYIKTEEDVPETSPININTPKNLSTSVLLSWDPVPKNGALISYDLTLQGPNEY 1195
DB	121 LNPCTTYEIKVAENAGAGVFDPLFQTAES-APGKVNLTVAYNASAVKLIWYLR 179	1137 TYTAQLYIKTEEDVPETSPININTPKNLSTSVLLSWDPVPKNGALISYDLTLQGPNEY 1196
QY	180 QPNKITSFKISVKHARSGIVKDVSRVEDILTGLPECN-RNSEFLWSTASPSPTLG 238	1196 SPITSDNYIILBELSPFTLYSPFAAARTKGLGPSSILFFYTDSPVLPAPPQNLTLINCT 1255
DB	181 QPNKITSFKISVKHARSGIVKDVSRVEDILTGLPECN-RNSEFLWSTASPSPTLG 240	1197 SPITSDNYIILBELSPFTLYSPFAAARTKGLGPSSILFFYTDSPVLPAPPQNLTLINCT 1256
QY	239 RVTPPSRTTHSSSTLTQNEISSVYWKPIPSFVVTHLRPYTYLFEVSAATTEAGYIDSTIV 298	1256 SDFVWLKWSPLPGIIVKVSFKIHEHETDITYKNIISGFKTEAKLVGLEPVSTYSIRV 1315
DB	241 RVTPPSRTTHSSSTLTQNEISSVYWKPIPSFVVTHLRPYTYLFEVSAATTEAGYIDSTIV 300	1257 SDFVWLKWSPLPGIIVKVSFKIHEHETDITYKNIISGFKTEAKLVGLEPVSTYSIRV 1316
QY	299 RTPESVPEGPQNCVTGNITGKSFSLWDPPPTIVTCKFSRVVLYGPGSRIILDNSTKDLK 358	1316 SAFTKVGNGNQFSNVVKFTTQESVDPVQNMOCMATSWQSVLVKMDPPKKANGIITQYMW 1375
DB	301 RTPESVPEGPQNCVTGNITGKSFSLWDPPPTIVTCKFSRVVLYGPGSRIILDNSTKDLK 360	1317 SAFTKVGNGNQFSNVVKFTTQESVDPVQNMOCMATSWQSVLVKMDPPKKANGIITQYMW 1376
QY	359 FAFNTLTPFTMYDYIAAETSAGTGPKSNISVTPPDVPGAVFDLQAEVESTQVRIWK 418	1376 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGBDESTCHVSTLPTVPVPTN 1435
DB	361 FAFNTLTPFTMYDYIAAETSAGTGPKSNISVTPPDVPGAVFDLQAEVESTQVRIWK 420	1377 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGBDESTCHVSTLPTVPVPTN 1436
QY	419 KRPQNGIINOYKVKVLPETGIIILENTLLTGNEVINDPMAPEIINIVEPMVGLYEGSA 478	1436 IAPSDVQSTSATLTWIRPDTILGYFQNYKIITTLRAQCKEWESECEVQKIQYLYEAH 1495
DB	421 KRPQNGIINOYKVKVLPETGIIILENTLLTGNEVINDPMAPEIINIVEPMVGLYEGSA 479	1437 IAPSDVQSTSATLTWIRPDTILGYFQNYKIITTLRAQCKEWESECEVQKIQYLYEAH 1496
QY	479 EMSSDLHSLATFIYNSHPDKNPPARNRAEDQTSFVVVTRNOYITDIAAEQLSVIRRLVP 538	1496 L7EETVYGLKFRWRFRVAASTNAGYGNASNIWSTKTLPGPPDPGPNNVHVATSPFSI 1555
DB	480 EMSSDLHSLATFIYNSHPDKNPPARNRAEDQTSFVVVTRNOYITDIAAEQLSVIRRLVP 539	1497 L7EETVYGLKFRWRFRVAASTNAGYGNASNIWSTKTLPGPPDPGPNNVHVATSPFSI 1556
QY	539 FTEHMSVSAFTMGSGPPTVLSVTRQVPSIKIINYNKISSSSILLYWDPPEYNGK 598	1556 S1SWEPAVITGPTCYLIDVKSVDNDEFNISPKNEENKTIIEIKDLEIFTRYSVVITAF 1615
DB	540 FTEHMSVSAFTMGSGPPTVLSVTRQVPSIKIINYNKISSSSILLYWDPPEYNGK 599	1557 S1SWEPAVITGPTCYLIDVKSVDNDEFNISPKNEENKTIIEIKDLEIFTRYSVVITAF 1616
QY	599 ITHYTIYAMELDTNRAFOITIDNSFLIT--GLKYYTKYKMRVAASTHDSGSSLENDI 656	1616 TGNISAAVVEGSSAEMIVTTLESAPKOPPPNNMTFKIPDEVTKQLTLPSPQNGNIQ 1675
DB	600 ITHYTIYAMELDTNRAFOITIDNSFLIT--GLKYYTKYKMRVAASTHDSGSSLENDI 659	1617 TGNISAAVVEGSSAEMIVTTLESAPKOPPPNNMTFKIPDEVTKQLTLPSPQNGNIQ 1676
QY	657 FVRTSEDEPSSPDVEVDVTADEIRLKWSPPEKPNGLIIIAVEVLYKKNIDTLYMKNST 716	1676 VYQALVYRDDPTAVQIHNLSIIQKTNFTVIAWLEGLKGGHTYNSVYAVNSAGAGPKVP 1735
DB	660 FVRTSEDEPSSPDVEVDVTADEIRLKWSPPEKPNGLIIIAVEVLYKKNIDTLYMKNST 719	1677 VYQALVYRDDPTAVQIHNLSIIQKTNFTVIAWLEGLKGGHTYNSVYAVNSAGAGPKVP 1736
QY	717 TDIIILNLRPHLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI 776	1736 MRITWIDIKAPARPKTPTPIYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQVLATET 1795
DB	720 TDIIILNLRPHLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI 779	1737 MRITWIDIKAPARPKTPTPIYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQVLATET 1796
QY	777 ELSFLPPSPNGIISKYTYLXESNGNEERTINTTSLTONIKVLKKTQYIIEVSASTLK 836	1796 GAQHDGNTKWDAYFNKARPYFTNEGPNPCTGKTKFSGNEEIIYIIGADNACMIPGN 1855
DB	780 ELSFLPPSPNGIISKYTYLXESNGNEERTINTTSLTONIKVLKKTQYIIEVSASTLK 837	1797 GAQHDGNTKWDAYFNKARPYFTNEGPNPCTGKTKFSGNEEIIYIIGADNACMIPGN 1856
QY	837 GEGVRSAPISILTEEDAPSPDQFVKOLSGVTVKLSWQPPLEPNGLIYTVYVW-NR 895	1856 EDKICNGPLKPKQYLFKFRATNIMGFTDSYSDPVKTLGEGLSERTVEIILSVTLCLL 1915
DB	838 GEGVRSAPISILTEEDAPSPDQFVKOLSGVTVKLSWQPPLEPNGLIYTVYVWNR 897	1857 EDKICNGPLKPKQYLFKFRATNIMGFTDSYSDPVKTLGEGLSERTVEIILSVTLCLL 1916
QY	896 SSLKTNVTETSLSDLDYNVEYSAYVTASTRFGDGKTSNIIISFQTEGAPSPDKDV 955	1916 SIILLGTAIFAPARIRQKQGGTYSPODAEIIIDTKLKDQITVADLELKDRLTR--- 1972
DB	898 SSLKTNVTETSLSDLDYNVEYSAYVTASTRFGDGKTSNIIISFQTEGAPSPDKDV 956	1917 SIILLGTAIFAPARIRQKQGGTYSPODAEIIIDTKLKDQITVADLELKDRLTRLLS 1976
QY	956 YNANLSSSSIIILFWTPPSKNGIIQYYSYVYRNTSGTFMONTLHETLNDPDMVTSTII 1015	1973 -----PISKSGFLQHVBEELCTNNNLKFOEESPELPKFQDLSSDADLPNNRKNRFPN 2026
DB	957 YNANLSSSSIIILFWTPPSKNGIIQYYSYVYRNTSGTFMONTLHETLNDPDMVTSTII 1016	1977 YKSKIPISKSGFLQHVBEELCTNNNLKFOEESPELPKFQDLSSDADLPNNRKNRFPN 2036
		2027 IKPYNNNRVKLIADASVPGSDYNASYSGLVCNEFIATQGLPGTGVDFRWWVWETR 2086
		2037 IKPY--NNNRVKLIADASVPGSDYNASYSGLVCNEFIATQGLPGTGVDFRWWVWETR 2095
		2087 AKTLVMLTQCPEKGRIRCHQWPNEDKNPVTVEGDIVITKLMEDVQIDWTIRDLKIERHGD 2146

Db 2096 AKTLVMTQCFKGRIRCHQYWPEDNKPVTVEGDIVITTKLMEDVQIDWTIRDLKIERHGD 2155
Qy 2147 CMTVRCNCTANPEHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVFIALD 2206
Db 2156 CMTVRCNCTANPEHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVFIALD 2215
Qy 2207 HLTOHINDHDFVDIYGLVAELSRMCMVONLAQYIFLHQCILDLLSNKGSNQPICFVNY 2266
Db 2216 HLTOHINDHDFVDIYGLVAELSRMCMVONLAQYIFLHQCILDLLSNKGSNQPICFVNY 2275
Qy 2267 SALQKMSLDAMEGDVELEWEEETM 2291
Db 2276 SALQKMSLDAMEGDVELEWEEETM 2300
RESULT 5
ADH41625
ID ADH41625 standard; protein; 2300 AA.
XX
AC ADH41625;
XX
DT 25-MAR-2004 (first entry)
XX
DE Novel human protein NOV15g.
XX
KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
KW antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX
OS Homo sapiens.
XX
PN WO2003102159-A2.
XX
XX 11-DEC-2003.
XX
PF 04-JUN-2003; 2003WO-US017573.
XX
PR 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386864P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 07-JUN-2002; 2002US-0387083P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387866P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.

PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390066P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 19-JUN-2002; 2002US-0391728P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.
XX
XX (CURA-) CURAGEN CORP.
XX
PA Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;
PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JH, Ji W, Kekuda R;
PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacIachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;
XX
XX WPI; 2004-053467/05.
XX N-PSDB; ADH41624.
XX
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.
XX
XX 'Claim 2; SEQ ID NO 178; 1503pp; English.
XX
XX The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the polypeptide of the invention.
XX
SQ Sequence 2300 AA;
Query Match 98.5%; Score 11807.5; DB 8; Length 2300;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2276; Conservative 5; Mismatches 5; Indels 19; Gaps 9;
Qy 1 MDFLIFILLFIFTGETQVDVSNVPGTRDYITISSITTYTSPVTRIVTNVTFPGPPV 60
Db 1 MDFLIFILLFIFTGETQVDVSNVPGTRDYITISSITTYTSPVTRIVTNVTFPGPPV 60

QY 61 FLA GER VGSAGILLSWNTPPNPNRGRISYIVKKEVCWPMQVYTOVRSPKDSLEVLLTN 120
DB 61 FLA GER VGSAGILLSWNTPPNPNRGRISYIVKKEVCWPMQVYTOVRSPKDSLEVLLTN 120
QY 121 LNPCTTYEIKVAEENSAGICVSPDPPLPQTAES-APGKVVNLTVYEAASAVKLIWYLP 179
DB 121 LNPCTTYEIKVAEENSAGICVSPDPPLPQTAESAPGKVVNLTVYEAASAVKLIWYLP 180
QY 180 QPKGKITSFKISVKHARSIGVVRDVSIRVEDILITGLKPECN-ENSESFLMSTASPTLG 238
DB 181 QPKGKITSFKISVKHARSIGVVRDVSIRVEDILITGLKPECNVENSESFLMSTASPTLG 240
QY 239 RVTPPRTTHSSSTLQNEISSYWKBPISFWVTHLRPYTYILFEVSAATTEAGYIDSTIV 298
DB 241 RVTPPRTTHSSSTLQNEISSYWKBPISFWVTHLRPYTYILFEVSAATTEAGYIDSTIV 300
QY 299 RTPESVPEGPQNCVTGNTGKSFSLMDPPTIVTGKFSVRVELYGPGRILDNSTKDLK 358
DB 301 RTPESVPEGPQNCVTGNTGKSFSLMDPPTIVTGKFSVRVELYGPGRILDNSTKDLK 360
QY 359 FAFNLTPFTMYDYIAAETSAGTGPKNISVFTPPDPVGA VFDLQLAEVESQVRITWK 418
DB 361 FAFNLTPFTMYDYIAAETSAGTGPKNISVFTPPDPVGA VFDLQLAEVESQVRITWK 420
QY 419 KRPQNGIINQYRVKVLVPETGIILENLTLTGNEVINDPMAPEIINIVEPMVGLYEGSA 478
DB 421 KRPQNGIINQYRVKVLVPETGIILENLTLTGNE-INDPMAPEIINIVEPMVGLYEGSA 479
QY 479 EMSDDLHSLATFIYNSHDPKNFARNRAEDQTSVVTTRNOYITDAAEQLSVIRLVP 538
DB 480 EMSDDLHSLATFIYNSHDPKNFARNRAEDQTSVVTTRNOYITDAAEQLSVIRLVP 539
QY 539 FTEHMLSVSAFTIMGEGPPTVLSVTRTQVPSSEIKIINYKNISSSILLYWDPEYPNGK 598
DB 540 FTEHMLSVSAFTIMGEGPPTVLSVTRTQVPSSEIKIINYKNISSSILLYWDPEYPNGK 599
QY 599 ITHYTYIAMELDTNRAFOITIDNSFLIT--GLKYYTKYKMRVAASHTHGPSSLSSENDI 656
DB 600 ITHYTYIAMELDTNRAFOITIDNSFLITIGLKKYTKYKMRVAASHTHGPSSLSSENDI 659
QY 657 FVTSDEDESSQDVEIDVTADEIRLKWSPKENGIIIAEVLVYKNIIDITLYMKTST 716
DB 660 FVTSDEDESSQDVEIDVTADEIRLKWSPKENGIIIAEVLVYKNIIDITLYMKTST 719
QY 717 TDIIILNLRPHLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSEGI 776
DB 720 TDIIILNLRPHLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSEGI 779
QY 777 ELSFLPPSPNGIIGKYTYILKESNGNEERTINTTSLQNIKVLKYYQYIIIEVSASTLK 836
DB 780 ELSFLPPSPNGIIGKYTYILKESNGNEERTINTTSLQNI--LKKYQYIIIEVSASTLK 837
QY 837 GEGVRSAPIISILTEEDAPSPDQFVKQSGVTVKLSWQPPLEPNGLIILYTYVYVW-NR 895
DB 838 GEGVRSAPIISILTEEDAPSPDQFVKQSGVTVKLSWQPPLEPNGLIILYTYVYVWRN 897
QY 896 SSLKTNVTSLELSDLDVNVESYSAVVTASTFRFGKTKGSNIISFQTPGAPSDPPKDV 955
DB 898 SSLKTNVTSLELSDLDVNVESYSAVVTASTFRFGKTKGSNIISFQTPGAPSDPPKDV 956
QY 956 YYANLSSSSIIILFWTPPSKENGIIQYYSVYRNTSGTFMGNFTLHBLTNDPDMNTVSTII 1015
DB 957 YYANLSSSSIIILFWTPPSKENGIIQYYSVYRNTSGTFMGNFTLHBLTNDPDMNTVSTII 1016
QY 1016 DKLTIFSYTFMLTASTSVGNKSSDIIIEVYTDQDIPGFGVNLVYESISSTAINVSW 1075
DB 1017 DKLTIFSYTFMLTASTSVGNKSSDIIIEVYTDQDIPGFGVNLVYESISSTAINVSW 1076
QY 1076 PPAQNGLYVYVSLIILOTPRHRPPLVYERSIYFDNLEKYTDVILKITSTEGKFS 1135
DB 1077 PPAQNGLYVYVSLIILOTPRHRPPLVYERSIYFDNLEKYTDVILKITSTEGKFS 1136

QY 1136 TYTAQLYIKTEEDVPETSPDIINTFKNLSSTSVLLSWDPVKNGALISVDLTLOGPNENY 1195
DB 1137 TYTAQLYIKTEEDVPETSPDIINTFKNLSSTSVLLSWDPVKVKNGALISVDLTLOGPNENY 1196
QY 1196 SPTTSDNYIIILBELSPFTLYSPFAAARTKGLGPSSILFPYTDDESVP LAPPNQLTINCT 1255
DB 1197 SPTTSDNYIIILBELSPFTLYSPFAAARTKGLGPSSILFPYTDDESVP LAPPNQLTINCT 1256
QY 1256 SDFVWLKWSPLPGGIVKVFYKFIHEHETDTIYKXNISGFKTEAKLVGLPVSYSIRV 1315
DB 1257 SDFVWLKWSPLPGGIVKVFYKFIHEHETDTIYKXNISGFKTEAKLVGLPVSYSIRV 1316
QY 1316 SAFTKVGNGNQSNNVKFTTQBSVPDVONMOCMATSWOSVLVKWDPKKKANGIIYQYV 1375
DB 1317 SAFTKVGNGNQSNNVKFTTQBSVPDVONMOCMATSWOSVLVKWDPKKKANGIIYQYV 1376
QY 1376 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGDESTCHVSTLPETVPSVPTN 1435
DB 1377 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGDESTCHVSTLPETVPSVPTN 1436
QY 1436 IAFSDVQSTSATLTWIRPDTIILGYFQNYKITTLQRAQCKEWESEBECVYQKIYLYEAH 1495
DB 1437 IAFSDVQSTSATLTWIRPDTIILGYFQNYKITTLQRAQCKEWESEBECVYQKIYLYEAH 1496
QY 1496 LTEETVYGLKXWRWRFQVAASNAGYGNASWISTKTLPGLPPDGPENVHVATSPFSI 1555
DB 1497 LTEETVYGLKXWRWRFQVAASNAGYGNASWISTKTLPGLPPDGPENVHVATSPFSI 1556
QY 1556 SJSWSBPAVITGPTCYLIDVKSVDNDEFNISPTKSNENKTIKOLEIFTRYSVVITAF 1615
DB 1557 SJSWSBPAVITGPTCYLIDVKSVDNDEFNISPTKSNENKTIKOLEIFTRYSVVITAF 1616
QY 1616 TGNISAAAYVEGSSAEMI VTTLESAPKOPNNMTFQKI PDEVTKFOLTLPSPQPNQNIQ 1675
DB 1617 TGNISAAAYVEGSSAEMI VTTLESAPKOPNNMTFQKI PDEVTKFOLTLPSPQPNQNIQ 1676
QY 1676 VYQALVYREDDPTAVQIHNLSIIQKNTNVIAMLEGLKGGHTYNI SVYAVNSAGAKVP 1735
DB 1677 VYQALVYREDDPTAVQIHNLSIIQKNTNVIAMLEGLKGGHTYNI SVYAVNSAGAKVP 1736
QY 1736 MRITWIDIKAPAPKTKPTPIYDATGKLVSTTITIRMPICYSDHGPKNQVVLATET 1795
DB 1737 MRITWIDIKAPAPKTKPTPIYDATGKLVSTTITIRMPICYSDHGPKNQVVLATET 1796
QY 1796 GAQHDGNVTKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMIPGN 1855
DB 1797 GAQHDGNVTKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMIPGN 1856
QY 1856 EDKICNGPLKPKKQVLFKPRATNIMGQPTDSYSDPVKTLGEGLSBRTVEIILSVTLCL 1915
DB 1857 EDKICNGPLKPKKQVLFKPRATNIMGQPTDSYSDPVKTLGEGLSBRTVEIILSVTLCL 1916
QY 1916 SIIILGTAIFAPARIQKQEGGTVSPODABIIDTKLKDOLITVADLEKDBRLTR--- 1972
DB 1917 SIIILGTAIFAPARIQKQEGGTVSPODABIIDTKLKDOLITVADLEKDBRLTRLLS 1976
QY 1973 -----PISKSGFLOHVEBELCTNNNLKFOEBSSELPKFLQDLSSSTDADLPWNAKRNFPN 2026
DB 1977 YRKSIPKISKSGFLOHVEBELCTNNNLKFOEBSSELPKFLQDLSSSTDADLPWNAKRNFPN 2036
QY 2027 IKPYNNNNRVKLIADASVPGSDYINASYISGYLCPNEPIATQGPLGTGDFWVWVETR 2086
DB 2037 IKPY-NNNRVKLIADASVPGSDYINASYISGYLCPNEPIATQGPLGTGDFWVWVETR 2095
QY 2087 AKTLVMTQCPKGRIRCHOYWPEDKNKPVTVFGDVIITKLMEDVQIDWIRDKIERHGD 2146
DB 2096 AKTLVMTQCPKGRIRCHOYWPEDKNKPVTVFGDVIITKLMEDVQIDWIRDKIERHGD 2155
QY 2147 CMTVQCNTFAMPEHGVSPENSAPLTHFVKLVASRAHDTTPMI VHSAGVGTGVFIAD 2206
DB 2156 CMTVQCNTFAMPEHGVSPENSAPLTHFVKLVASRAHDTTPMI VHSAGVGTGVFIAD 2215
QY 2207 HLTOHINDHDFVDIYGLVAELSRMCMVQNLQAQVIFLHQCILDLNLSKNSQPCFVNY 2266

/note="Potentially phosphorylated"	208
/note="Potentially phosphorylated"	249
/note="Potentially phosphorylated"	298
/note="Potentially phosphorylated"	300..383
/label= Fibronectin_type_III_domain	/note="Identified using HMMER_PPAM"
/note="Potentially glycosylated"	313
/note="Potentially phosphorylated"	328
/note="Potentially phosphorylated"	332
/note="Potentially phosphorylated"	349
/note="Potentially glycosylated"	350
/note="Potentially phosphorylated"	365
/note="Potentially phosphorylated"	384
/note="Potentially glycosylated"	389
/note="Potentially phosphorylated"	394..554
/label= Fibronectin_type_III_domain	/note="Identified using HMMER_PPAM"
/note="Potentially phosphorylated"	490
/note="Potentially phosphorylated"	511
/note="Potentially phosphorylated"	557
/note="Potentially phosphorylated"	565..651
/label= Fibronectin_type_III_domain	/note="Identified using HMMER_PPAM"
/note="Potentially phosphorylated"	575
/note="Potentially glycosylated"	607
/note="Potentially phosphorylated"	614
/note="Potentially phosphorylated"	623
/note="Potentially phosphorylated"	633
/note="Potentially phosphorylated"	649
/note="Potentially phosphorylated"	660
/note="Potentially phosphorylated"	661
/note="Potentially phosphorylated"	665..746
/label= Fibronectin_type_III_domain	/note="Identified using HMMER_PPAM"
/note="Potentially phosphorylated"	678
/note="Potentially phosphorylated"	687
/note="Potentially phosphorylated"	710
/note="Potentially phosphorylated"	713
/note="Potentially glycosylated"	713

FT	Modified-site	715	/note= "Potentially phosphorylated"	QY	61	FLAGRVGSAGILLSWNTPPNPNGRIISYIVKYKVCWPMQTVYTVQVRSPDSLEVLTTN	120
FT	Modified-site	731	/note= "Potentially glycosylated"	Db	56	FLAGRVGSAGILLSWNTPPNPNGRIISYIVKYKVCWPMQTVYTVQVRSPDSLEVLTTN	115
FT	Modified-site	733	/note= "Potentially glycosylated"	QY	121	LNPGTTYEIKVAAENSAGIGVFDPLFOTAESAPGKVNLTVYANASAVKLIWYLPQ	180
FT	Modified-site	751	/note= "Potentially phosphorylated"	Db	116	LNPGTTYEIKVAAENSAGIGVFDPLFOTAESAPGKVNLTVYANASAVKLIWYLPQ	175
FT	Modified-site	757	/note= "Potentially phosphorylated"	QY	181	PNGKITSPKISVKHARSGIVVVDVIRVEDILTGLPECNENSESEFLWSTASPSPTLGRV	240
FT	Domain	759..842	/label= "Fibronectin type III domain /note= "Identified using HMMER_Pfam"	Db	176	PNGKITSPKISVKHARSGIVVVDVIRVEDILTGLPECNENSESEFLWSTASPSPTLGRV	235
FT	Modified-site	761	/note= "Potentially phosphorylated"	QY	241	TPPSRTTHSSSTLTONEISSWKKEPISFVWTHLRPTTYLFEVSAATTAGYIDSTIVRT	300
FT	Modified-site	765	/note= "Potentially phosphorylated"	Db	236	TPPSRTTHSSSTLTONEISSWKKEPISFVWTHLRPTTYLFEVSAATTAGYIDSTIVRT	295
FT	Modified-site	767	/note= "Potentially glycosylated"	QY	301	PESVPEGPPQNCVTGNITGKSPSILWDPPTIVTGFSYRVLYGPGS- -GRILLDNSTKDLK	358
FT	Modified-site	770	/note= "Potentially phosphorylated"	Db	296	PESVPEGPPQNCVTGNITGKSPSILWDPPTIVTGFSYRVLYGPGS- -GRILLDNSTKDLK	355
FT	Modified-site	772	/note= "Potentially glycosylated"	QY	359	PAFTNLTPPTMYDVVYIAAETSAGTGPKGNISVFTPPDVEGAVFDLQLAESTQVRAITWK	418
FT	Modified-site	809	/note= "Potentially phosphorylated"	Db	356	PAFTNLTPPTMYDVVYIAAETSAGTGPKGNISVFTPPDVEGAVFDLQLAESTQVRAITWK	415
FT	Modified-site	824	/note= "Potentially glycosylated"	QY	419	KRPQNGIINQYRVKVLVPETGIIILENTLLTGNEYINDPMAPEIYVIVPEWGLYEGSA	478
FT	Modified-site	834	/note= "Potentially phosphorylated"	Db	416	KRPQNGIINQYRVKVLVPETGIIILENTLLTGNE- -INDPMAPEIYVIVPEWGLYEGSA	474
FT	Modified-site	849	/note= "Potentially phosphorylated"	QY	479	EMSSDLHSLATPIYNSHPDKNFPARNRAEDQTSVVTTRNOYITDIAAQLSVIRRLVP	538
FT	Domain	854..936	/label= "Fibronectin type III domain /note= "Identified using HMMER_Pfam"	Db	475	EMSSDLHSLATPIYNSHPDKNFPARNRAEDQTSVVTTRNOYITDIAAQLSVIRRLVP	534
FT	Modified-site	862	/note= "Potentially phosphorylated"	QY	539	FTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINYKNISSSSILLYWMPPEYPNGK	598
FT	Modified-site	870	/note= "Potentially phosphorylated"	Db	535	FTEHMSVSAFTIMGEGPPTVLSVTRQOVPSIKIINYKNISSSSILLYWMPPEYPNGK	594
FT	Modified-site	895	/note= "Potentially glycosylated"	QY	599	ITHYTIYAMELDTNRAFOITITDINSFLITG- -LKYTKYKQVAASTHDESSISEEN	654
FT	Modified-site	898	/note= "Potentially glycosylated"	Db	595	ITHYTIYAMELDTNRAFOITITDINSFLITG- -LKYTKYKQVAASTHDESSISEEN	654
FT	Modified-site	903	/note= "Potentially phosphorylated"	QY	655	DIFVTSDEPSSPDQVEVDVTADEIRLKWSPPEKNGIILIAEVLKNIDTLVMKNT	714
FT	Modified-site	907	/note= "Potentially glycosylated"	Db	655	DIFVTSDEPSSPDQVEVDVTADEIRLKWSPPEKNGIILIAEVLKNIDTLVMKNT	714
FT	Modified-site	912	/note= "Potentially phosphorylated"	QY	715	STTDIILNLRPHLYNISVRSYTRFGHGNQVSSLSVRSVTSVTPDSPAENITYKNISG	774
FT	Modified-site	927	/note= "Potentially phosphorylated"	Db	715	STTDIILNLRPHLYNISVRSYTRFGHGNQVSSLSVRSVTSVTPDSPAENITYKNISG	774
FT	Domain	948..1040	/label= "Fibronectin type III domain /note= "Identified using HMMER_Pfam"	QY	775	EIELSFLPPSPNGIISKYTIYLRKSGNEERTINTSLTQNIKGLKKYQTVIIEVSAST	834
FT	Modified-site	959	/note= "Potentially glycosylated"	Db	775	EIELSFLPPSPNGIISKYTIYLRKSGNEERTINTSLTQNIKGLKKYQTVIIEVSAST	834
FT	Modified-site	988	/note= "Potentially glycosylated"	QY	835	LKGEGRSAPISILTEEDAPSPDQFVSVKQSGVTVKLSWQPPLENGIILYYTVYVW	893
FT	Modified-site	996	/note= "Potentially glycosylated"	Db	835	LKGEGRSAPISILTEEDAPSPDQFVSVKQSGVTVKLSWQPPLENGIILYYTVYVW	894
FT	Modified-site	998	/note= "Potentially glycosylated"	QY	894	NRSSLKTINVTETSLSDLDYNYEYSAYVASTRFGDKTGSNIISFOTPEGASDPDK	953
FT	Modified-site	1008	/note= "Potentially phosphorylated"	Db	895	NRSSLKTINVTETSLSDLDYNYEYSAYVASTRFGDKTGSNIISFOTPEGASDPDK	953
FT	Modified-site	1013	/note= "Potentially glycosylated"	QY	954	DVYANLSSSSIIILFWTPPSKNGIIOYYSVYRNTSGTFMONTHELNDNDNTVST	1013
FT	Modified-site	1013	/note= "Potentially glycosylated"	Db	954	DVYANLSSSSIIILFWTPPSKNGIIOYYSVYRNTSGTFMONTHELNDNDNTVST	1013
FT	Query Match		98.4%; Score 11797; DB 6; Length 2299;	QY	1014	IIDKLTIYSYTFWLTASTSVGNKSSDIIEVYDQDIPGCFVGNLTYESSISAINVS	1073
FT	Best Local Similarity		98.5%; Pred. No. 0;	Db	1014	IIDKLTIYSYTFWLTASTSVGNKSSDIIEVYDQDIPGCFVGNLTYESSISAINVS	1073
FT	Matches 2272; Conservative		5; Mismatches 5; Indels 24; Gaps 8;	QY	1074	WPPAQNGLVFPYVSLILQOTPHRVPLVYERSIVFDNLEKTDYDKLITPTEKGF	1133
FT	1 MDPLIFLLFLTGTSFQVDSNVVPGTRYDYITISISTYTSVPTRIYTPNVTKGPPV		60	Db	1074	WPPAQNGLVFPYVSLILQOTPHRVPLVYERSIVFDNLEKTDYDKLITPTEKGF	1133
FT	1 MDPLIFLLFLTGTSFQVDSNVVPGTRYDYITISISTYTSVPTRIYTPNVTKGPPV		55	QY	1134	SDTYTAQLYIKTEEDVPETSPINTFKNLSSVLLSWDPPVVKPNGAIISYDLTLQGPNE	1193

|||||
1134 SDTYTAQYIKTEEDVPETSPINTPKNLSSVLLSWDPVPKNGALISYDLTLQGPNE 1193
|||||
1194 NYSFITSNDYIILEELSPPTLYSFFFAAARTRKGLGPSSILFFYTDESVP LAPQMLTLIN 1253
|||||
1194 NYSFITSNDYIILEELSPPTLYSFFFAAARTRKGLGPSSILFFYTDESVP LAPQMLTLIN 1253
|||||
1254 CTSDFVWLKWSPLPGGIVKVYSPKIHHEHETDTIYKNI SGFKTEAKLVGLEPSTYSI 1313
|||||
1254 CTSDFVWLKWSPLPGGIVKVYSPKIHHEHETDTIYKNI SGFKTEAKLVGLEPSTYSI 1313
|||||
1314 RVSAFTKVGNGQFNSNVVFTTQESVPDVQNMCMATSWQSVLVKWDPPKKANGIITQY 1373
|||||
1314 RVSAFTKVGNGQFNSNVVFTTQESVPDVQNMCMATSWQSVLVKWDPPKKANGIITQY 1373
|||||
1374 MVTVERNSTKVPQDMYTFIKLLANTS YVFKVRASTASAGEDESTCHVSTLPETVPSPV 1433
|||||
1374 MVTVERNSTKVPQDMYTFIKLLANTS YVFKVRASTASAGEDESTCHVSTLPETVPSPV 1433
|||||
1434 TNAFSDVOSTATLTWIRPDITLGYFQNYKITTQRAOKCKEWESEECVEYKQIQLYE 1493
|||||
1434 TNAFSDVOSTATLTWIRPDITLGYFQNYKITTQRAOKCKEWESEECVEYKQIQLYE 1493
|||||
1494 AHLTEETVYGLKKFRWYRQVAASSTNAGYGNASNMISTKTLPQPPDPPENHVVATSPF 1553
|||||
1494 AHLTEETVYGLKKFRWYRQVAASSTNAGYGNASNMISTKTLPQPPDPPENHVVATSPF 1553
|||||
1554 SISISWSEPAVITGPTCYLIDKVSVDNDFNIFSKSNEENKTIEIKOLEIFTRYSVVIT 1613
|||||
1554 SISISWSEPAVITGPTCYLIDKVSVDNDFNIFSKSNEENKTIEIKOLEIFTRYSVVIT 1613
|||||
1614 AFTGNLSAAVVEGKSAEMI VTTLESAPKDPNNMTFOKIPDRVTKFQLTLPSPQNGN 1673
|||||
1614 AFTGNLSAAVVEGKSAEMI VTTLESAPKDPNNMTFOKIPDRVTKFQLTLPSPQNGN 1673
|||||
1674 IQVYQALVYREDPTAVQIHNLSIIQKTNFTFIAMLEGLKGHTYNI SYVAVNSAGAGPK 1733
|||||
1674 IQVYQALVYREDPTAVQIHNLSIIQKTNFTFIAMLEGLKGHTYNI SYVAVNSAGAGPK 1733
|||||
1734 VPMRITMDIKAPARPKTPTIYDATGKLLVSTTITIRMPICYSDHDPKINQVVLAT 1793
|||||
1734 VPMRITMDIKAPARPKTPTIYDATGKLLVSTTITIRMPICYSDHDPKINQVVLAT 1793
|||||
1794 ETGAOHGDNVTWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMP 1853
|||||
1794 ETGAOHGDNVTWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIGADNACMP 1853
|||||
1854 GNEDKICNGPLKPKQYLPKFRATNIMGFTSDSDYSDPVKTLGEGLSERTVEIILSVTL 1913
|||||
1854 GNEDKICNGPLKPKQYLPKFRATNIMGFTSDSDYSDPVKTLGEGLSERTVEIILSVTL 1913
|||||
1914 ILSIILLGTAIFAFARIRQKQEGGYSPQDAEIIDTKLQDLITVADLELKDERLTR- 1972
|||||
1914 ILSIILLGTAIFAFARIRQKQEGGYSPQDAEIIDTKLQDLITVADLELKDERLTR 1973
|||||
1973 -----PISKKSFLQHVHEELCTNNNLKFOREFSELPKFLODLSSTDADLPWNRKNRF 2024
|||||
1974 LSYRKSIFPISKKSFLQHVHEELCTNNNLKFOREFSELPKFLODLSSTDADLPWNRKNRF 2033
|||||
2025 PNIKPYNNNRVLKIDASVPGSDYNASVIGYLCNPNEFIAQTGSLPGTVGDFWRWVWE 2084
|||||
2034 PNIKPY-NNNRVLKIDASVPGSDYNASVIGYLCNPNEFIAQTGSLPGTVGDFWRWVWE 2092
|||||
2085 TRAKTLVMTLQCEKGRIRCHQFWPBDNKPVTVPFGDIVITKLMEDVQIDMTWTRDLKIEH 2144
|||||
2093 TRAKTLVMTLQCEKGRIRCHQFWPBDNKPVTVPFGDIVITKLMEDVQIDMTWTRDLKIEH 2152
|||||
2145 GDCMTVRCNQFTTAWPBEHVPENSAPLIHFVKLVRSRAHDTTPMIVHCSAGVGRGTGFTA 2204
|||||
2153 GDCMTVRCNQFTTAWPBEHVPENSAPLIHFVKLVRSRAHDTTPMIVHCSAGVGRGTGFTA 2212
|||||
2205 LDHLTQHINDHFDVDIYGLVAELSRRCMCMQVNLQAYIFLHQICILDLLSNKGSNOPICFV 2264
|||||

Db 2213 LDHLTQHINDHFDVDIYGLVAELSRRCMCMQVNLQAYIFLHQICILDLLSNKGSNOPICFV 2272
QY 2265 NYSALQKQWSDLDAMEGDVELEWEETM 2291
Db 2273 NYSALQKQWSDLDAMEGDVELEWEETM 2299
RESULT 7
AAO18736
ID AAO18736 standard; protein; 2281 AA.
XX
AC AAO18736;
XX
DT 24-OCT-2002 (first entry)
XX
DE Human NOV2a protein.
XX
KW Human; NOVX; autoimmune disease; cancer; infection; inflammatory disease;
KW storage disorder; muscle disorder; neurodegenerative disorder; nontropic;
KW developmental defect; neuroprotective; antiparkinsonian; hypotensive;
KW hypertensive; haemostatic; cardiac; antitumoral; dermatological;
KW immunosuppressive; antiinflammatory; virucide; antibacterial; anti-HIV;
KW antiparasitic; antiallergic; antidiabetic; immunomodulator; antipsoriatic;
KW nephrotropic; anorectic; antidiabetic; immunomodulator; antipsoriatic;
KW antinfertility; antitumor; antidiabetic; immunomodulator; antipsoriatic;
KW tranquilizer; analgesic.
XX
OS Homo sapiens.
XX
PN WO200257450-A2.
XX
PD 25-JUL-2002.
XX
PF 29-NOV-2001; 2001WO-US048922.
XX
PR 29-NOV-2000; 2000US-0253834P.
PR 30-NOV-2000; 2000US-0250926P.
PR 25-JAN-2001; 2001US-0264180P.
PR 20-AUG-2001; 2001US-0313658P.
PR 05-OCT-2001; 2001US-0327458P.
PR 28-NOV-2001; 2001US-00327456.
XX
(CURA-) CURAGEN CORP.
XX
PI Edinger S, Macdougall JR, Millet I, Ellemann K, Stone DJ, Burgess CE;
PI Gerlach V, Grosse WM, Alsobrook JP, Lepley DM, Rieger D, Mishra V;
PI Casman SJ, Spytek KA, Boldog FL, Li L, Padigaru M, Mishra V;
PI Patturajan M, Shenoy S, Rastelli L, Tchernev VT, Vernet CAM;
PI Zerhusen BD, Malyankar UM, Guo X, Miller CE, Gangolli EA;
XX
WPI; 2002-590741/63.
DR N-PSDB; ABO6281.
XX
PT Novel isolated polypeptide, designated NOVX, useful for treating or
PT preventing in NOVX-associated disorders e.g. cardiomyopathy,
PT atherosclerosis, diabetes, cancer, allergy, asthma, Crohn's disease.
XX
PS Claim 1; Page 25-26; 353pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC several novel human proteins, designated NOVX. These can be used in the
CC treatment of, amongst others, cancers, autoimmune diseases, infections,
CC inflammatory diseases, storage disorders, muscle disorders, the present
CC neurodegenerative diseases and developmental defects. The present
CC sequence is a protein of the invention
XX
SQ Sequence 2281 AA;
Query Match 93.4%; Score 11191; DB 5; Length 2281;
Best Local Similarity 94.5%; Pred. No. 0;
Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

QY 1 MDPLILFLFICTSETQVDSNVVPGTRVDITISSITSTYTSPVTRI VTPNVTKBGPV 60
Db 1 MDPLILFLFICTSETQVDSNVVPGTRVDITISSITSTYTSPVTRI GASN--EPGPPV 58
QY 61 FLA GERVGSAGILLSNWTPNPNNGRIISYIVKYKEVCPNMQTVYTVQRKPSLEVLN 120
Db 59 FLA GERVGSAGILLSNWTPNPNNGRIISYIVKYKEVCPNMQTVYTVQRKPSLEVLN 118
QY 121 LNFPTTYEIKVAENAGAGVFPDPLFQTAESAPGKVNLVVEAYNASAVKLIWLPQ 180
Db 119 LNFPTTYEIKVAENAGAGVFPDPLFQTAESAPGKVNLVVEAYNASAVKLIWLPQ 177
QY 181 PNCKITSPKSVKHSAGIIVKDVSVRVEDILTKLP-ECNENSESFLMSTASPTLGR 239
Db 178 KK-KITSFKLSVXHSRGIIVKDVSVRVEDILTKLP-ECNENSESFLMSTASPTLGR 236
QY 240 VTPPSRTTHSSSTLTQNEISSVMKEPISFVVTHLRPVTVTLFVSAATTEAGYIDSTIVR 299
Db 237 VTPPSRTTHSSSTLTQNEISSV-KEPISFVVTHLRPVTVTLFVSAATTEAGYIDSTIVR 295
QY 300 TPESVPEGPQNCVTGNITGKSPSILWDPPITVIGKFSYRVELYGPS-GRILDNSTKDLK 358
Db 296 TPESVPEGPQNCVTGNITGKSPSILWDPPITVIGKFSYRVELYGPS-GRILDNSTKDLK 355
QY 359 PAFNTLTPPTMYDVYIAAETSAGTGPKNISVETPDVPGAVDLOLAEVSTQVRIWK 418
Db 356 PAFNTLTPPTMYDVYIAAETSAGTGPKNISVETPDVPGAVDLOLAEVSTQVRIWK 415
QY 419 KRPQNGIINOYRKVLVPTGIIILENTLLTGNEVINDPMAPEIUNI VEPWGLYEGSA 478
Db 416 KRPQNGIINOYRKVLVPTGIIILENTLLTGNE-INDPMAPEIUNI VEPWGLYEGSA 474
QY 479 EMSDDLHSLATFIYNSHPDKNFPARNRAEDQTSFVVTTRNQYITDIAAEQLSVIRRLVP 538
Db 475 EMSDDLHSLATFIYNSHPDKNFPARNRAEDQTSFVVTTRNQYITDIAAEQLSVIRRLR 534
QY 539 FTHMISVAGFTMGSGPPTVLVRTRQQVPSSIKIINIK--NISSSILLYWDPEYEN 596
Db 535 FWAETMGFSRYTIMSS-----ASRDLNLTSPGSLAQNFRVTHVTITEVFLHWDPPD--P 586
QY 597 GKITHYTIWAMELDT-NRAQIITIDN-SPLITGLAKYTKYKMRVAASTHDGSSISEEN 654
Db 587 VFPHYLLITLOVENQSKSILKLTNSLVLGLKYYTKYKMRVAASTHVGSSISEEN 646
QY 655 DIFVRTSEDEPSSQDVEIDVTADEIRLKWSPPEKPGNGIIIAEVLVKNIDTLVKN 714
Db 647 DIFVRTSEDEPSSQDVEIDVTADEIRLKWSPPEKPGNGIIIAEVLVKNIDTLVKN 706
QY 715 STTDIILRLNRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISG 774
Db 707 STTDIILRLNRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISG 766
QY 775 EIELSLPSSPNGIILKYYTKYKMRVAASTHDGSSISEEN 834
Db 767 EIELSLPSSPNGIILKYYTKYKMRVAASTHDGSSISEEN 826
QY 835 LKGEGRVRSAPISILTBEDAPSPQDFSVKQLSGVTVKLSWQPLBNGIILYTVYVWN 894
Db 827 LKGEGRVRSAPISILTBEDAPSPQDFSVKQLSGVTVKLSWQPLBNGIILYTVYVWN 885
QY 895 RSSLKNTINVTETSLSDLYNVEYSAYVTASTRFGDGKGTGNSIISFQTPPEGAPSPK 954
Db 886 RSSLKNTINVTETSLSDLYNVEYSAYVTASTRFGDGKGTGNSIISFQTPPEG-PSDPPK 944
QY 955 VYIANLSSSSILLFWTPSPKNGIIQYYSVYRNTSGTFMONTLHELNDPDMVTSTI 1014
Db 945 VYIANLSSSSILLFWTPSPKNGIIQYYSVYRNTSGTFMONTLHELNDPDMVTSTI 1004
QY 1015 IDKLTIFSYTYFWLTASTSVGNKSSDIIIEVYTDQDIPGFGVGNLTYESISTALNVSM 1074
Db 1005 IDKLTIFSYTYFWLTASTSVGNKSSDIIIEVYTDQDIPGFGVGNLTYESISTALNVSM 1064
QY 1075 VPPAQPNGLVFFYVVSILQOTPRHVRPPLVTVRSIYFONLEKYTDYILKITPSTBKGS 1134

Db 1065 VPPAQPNGLVFFYVVSILQOTPRHVRPPLVTVRSIYFONLEKYTDYILKITPSTBKGS 1124
QY 1135 DTYTAQLYIKTEBEDVPETSPIIINTFNKLSSTSVLLSWDPPVPNGAIISYDLTLQGNEN 1194
Db 1125 DTYTAQLYIKTEBEDVPETSPIIINTFNKLSSTSVLLSWDPPVPNGAIISYDLTLQGNEN 1184
QY 1195 YSFITSDNYVIIIEELSPFTLYSFPFAAARTRKGLGSSIIFFYTDSEVPLAPPQNLILNC 1254
Db 1185 YSFITSDNYVIIIEELSPFTLYSFPFAAARTRKGLGSSIIFFYTDSEVPLAPPQNLILNC 1244
QY 1255 TSDFWLWKSPSPGPIVGVYSFKIHEHETDIYKNIISGFKTEAKLVGLEBVSYSIR 1314
Db 1245 TSDFWLWKSPSPGPIVGVYSFKIHEHETDIYKNIISGFKTEAKLVGLEBVSYSIR 1304
QY 1315 VSAFTKVGNGNFSNVVKFTTQESVDPVVQNNQCMATSWQSVLVKWDPPKANGIITQYM 1374
Db 1305 VSAFTKVGNGNFSNVVKFTTQESVDPVVQNNQCMATSWQSVLVKWDPPKANGIITQYM 1364
QY 1375 VTVERNSTKVSQDDHMYTPIKLLANTSYVFKVRASTASAGEDESTCHVSTLPTVSVPT 1434
Db 1365 VTVERNSTKVSQDDHMYTPIKLLANTSYVFKVRASTASAGEDESTCHVSTLPTVSVPT 1424
QY 1435 NIAPSDVQSTATLWIRPDTTILGYFQNYKITTLQRAQCKEWESECEYQIKIQLYEA 1494
Db 1425 NIAPSDVQSTATLWIRPDTTILGYFQNYKITTLQRAQCKEWESECEYQIKIQLYEA 1484
QY 1495 HLTETVYGLKFRWYRFOVAASSTNAGYGNASNMISTKTLPDPPDPPPNVHVATSPFS 1554
Db 1485 HLTETVYGLKFRWYRFOVAASSTNAGYGNASNMISTKTLPDPPDPPPNVHVATSPFS 1544
QY 1555 ISISWSEPAVITGPTCYLLIDVKSVDNDEFNISFIKSNBENKTEIHKDLBIFRYSVVITA 1614
Db 1545 ISISWSEPAVITGPTCYLLIDVKSVDNDEFNISFIKSNBENKTEIHKDLBIFRYSVVITA 1604
QY 1615 FTGNTSAAVVEGSSAEMIVTLESAPKDPNNMTFQKIPDEVTKFQLTFLPSPQNGNI 1674
Db 1605 FTGNTSAAVVEGSSAEMIVTLESAPKDPNNMTFQKIPDEVTKFQLTFLPSPQNGNI 1664
QY 1675 QYVQALVYREDDPTAVQIHNLSIIQKNTFVIALEGLKGGHTYNSVAVNSAGAGPKV 1734
Db 1665 QYVQALVYREDDPTAVQIHNLSIIQKNTFVIALEGLKGGHTYNSVAVNSAGAGPKV 1724
QY 1735 PMRITMDIKAPARPKTPPIYDATGKLIVTSTTITIRMPICIYSDDHGPIKRVQVLATE 1794
Db 1725 PMRITMDIKAPARPKTPPIYDATGKLIVTSTTITIRMPICIYSDDHGPIKRVQVLATE 1784
QY 1795 TGAQHDGNTWKYDAYFNKARPYFTNEGPNPCTEGTKFSGNEEIIYIIGADNACMIPG 1854
Db 1785 TGAQHDGNTWKYDAYFNKARPYFTNEGPNPCTEGTKFSGNEEIIYIIGADNACMIPG 1844
QY 1855 NEDKICNGPLKPKQVLPKFRATNIMGQFTSDYSDPVKTLGEGLSERTVEIILSVTLCI 1914
Db 1845 NEDKICNGPLKPKQVLPKFRATNIMGQFTSDYSDPVKTLGEGLSERTVEIILSVTLCI 1904
QY 1915 LSIILLGTAFAPARIRQKQEGGTVSPQDASIIIDTKLQDLITVADLEKDERLTRPI 1974
Db 1905 LSIILLGTAFAPARIRQKQEGGTVSPQDASIIIDTKLQDLITVADLEKDERLTRPI 1964
QY 1975 SKKSFLOHVEELCTNNLKFQBEFSELPKFLOQLSDTADLDPWNRKRNFPPIKPNNNN 2034
Db 1965 SKKSFLOHVEELCTNNLKFQBEFSELPKFLOQLSDTADLDPWNRKRNFPPIKPNNNN 2023
QY 2035 RVKLJADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWVWVETRAKTLVMLT 2094
Db 2024 RVKLJADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWVWVETRAKTLVMLT 2083
QY 2095 QCFEKGRIRCHOYWPEDNKPVTVFGDIIVITKLMEDVQIDWTIRDLKIERHGDCTVROCN 2154
Db 2084 QCFEKGRIRCHOYWPEDNKPVTVFGDIIVITKLMEDVQIDWTIRDLKIERHGDCTVROCN 2143
QY 2155 FTAWPEHGVPENSAPLILHPVKLVRASRAHDTTTPMIVHCSAGVGRGTGVTIALDHLTOHND 2214

Db 2144 FTAPWBGHPENSAPLIHFVKLVRAAGRAHDTTPMIVHCAGVGRGTGVFALDHLTQHIND 2203

Qy 2215 HDPVDIYGLVAELSRMCMQVLAQYIFLHQICLDLLSNKGSNQPICFVNTSALQKXMS 2274

Db 2204 HDPVDIYGLVAELSRMCMQVLAQYIFLHQICLDLLSNKGSNQPICFVNTSALQKXMS 2263

Qy 2275 LDAME-GDVELEWEETM 2291

Db 2264 LDAMEGGDVELEWEETM 2281

RESULT 8

ADH41619

ID ADH41619 standard; protein; 2281 AA.

XX

AC ADH41619;

XX

XX

25-MAR-2004 (first entry)

XX

XX Novel human protein NOV15d.

XX

KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;

KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;

KW antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;

KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;

KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;

KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;

KW tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX

OS Homo sapiens.

XX

XX

PN WO2003102159-A2.

XX

XX

PD 11-DEC-2003.

XX

XX

PF 04-JUN-2003; 2003WO-US017573.

XX

PR 04-JUN-2002; 2002US-0385490P.

PR 04-JUN-2002; 2002US-0385615P.

PR 04-JUN-2002; 2002US-0385755P.

PR 05-JUN-2002; 2002US-0386041P.

PR 06-JUN-2002; 2002US-0386355P.

PR 06-JUN-2002; 2002US-0386357P.

PR 06-JUN-2002; 2002US-0386447P.

PR 06-JUN-2002; 2002US-0386459P.

PR 06-JUN-2002; 2002US-0386465P.

PR 06-JUN-2002; 2002US-0386684P.

PR 07-JUN-2002; 2002US-0386701P.

PR 07-JUN-2002; 2002US-0386796P.

PR 07-JUN-2002; 2002US-0386931P.

PR 07-JUN-2002; 2002US-0387078P.

PR 07-JUN-2002; 2002US-0387081P.

PR 07-JUN-2002; 2002US-0387083P.

PR 10-JUN-2002; 2002US-0387429P.

PR 10-JUN-2002; 2002US-0387540P.

PR 10-JUN-2002; 2002US-0387866P.

PR 10-JUN-2002; 2002US-0387869P.

PR 11-JUN-2002; 2002US-0387610P.

PR 11-JUN-2002; 2002US-0387659P.

PR 11-JUN-2002; 2002US-0387668P.

PR 11-JUN-2002; 2002US-0387696P.

PR 11-JUN-2002; 2002US-0387859P.

PR 12-JUN-2002; 2002US-0387934P.

PR 12-JUN-2002; 2002US-0387960P.

PR 12-JUN-2002; 2002US-0388022P.

PR 12-JUN-2002; 2002US-0388096P.

PR 12-JUN-2002; 2002US-0388432P.

PR 12-JUN-2002; 2002US-0388479P.

PR 13-JUN-2002; 2002US-0389123P.

PR 14-JUN-2002; 2002US-0389120P.

PR 14-JUN-2002; 2002US-0389146P.

PR 17-JUN-2002; 2002US-0389742P.

PR 18-JUN-2002; 2002US-0389604P.

PR 18-JUN-2002; 2002US-0389884P.

PR 19-JUN-2002; 2002US-0390006P.

PR 19-JUN-2002; 2002US-0390144P.

PR 19-JUN-2002; 2002US-0390209P.

PR 25-JUN-2002; 2002US-0391726P.

PR 06-AUG-2002; 2002US-0401628P.

PR 09-AUG-2002; 2002US-0402268P.

PR 12-AUG-2002; 2002US-0402822P.

PR 13-AUG-2002; 2002US-0403458P.

PR 15-AUG-2002; 2002US-0403617P.

PR 15-AUG-2002; 2002US-0403732P.

PR 26-AUG-2002; 2002US-0406182P.

PR 12-SEP-2002; 2002US-0410085P.

PR 13-SEP-2002; 2002US-0410505P.

PR 23-SEP-2002; 2002US-0412955P.

PR 30-SEP-2002; 2002US-0415195P.

PR 23-OCT-2002; 2002US-0420627P.

PR 23-OCT-2002; 2002US-0420718P.

PR 24-OCT-2002; 2002US-0420852P.

PR 31-OCT-2002; 2002US-0422750P.

PR 01-NOV-2002; 2002US-0423095P.

PR 05-NOV-2002; 2002US-0423748P.

XX

XX (CURA-) CURAGEN CORP.

PA

XX

PI Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;

PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;

PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;

PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;

PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;

PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;

PI Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;

PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;

PI Wolenc AR, Zhong M, Zhong H;

XX

XX WPI: 2004-053467/05.

DR N-PSDB; ADH41618.

XX

XX

PT New NOVX polypeptides and nucleic acid molecules useful for preventing or

PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,

PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

PT pharmacogenomics.

XX

PS Claim 2; SEQ ID NO 172; 1503pp; English.

XX

XX The invention relates to 566 new isolated human polypeptides and their

XX encoding genes, sequences that are at least 95% identical to these or

XX sequences comprising one or more conservative substitutions in these. The

XX polypeptide, polynucleotide and antibodies against the polypeptides are

XX useful in diagnosing, treating or preventing NOVX-associated disorders,

XX e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,

XX diabetes, AIDS, multiple sclerosis, graft-versus-host disease,

XX Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.

XX The nucleic acids are further used as hybridization probes, in chromosome

XX mapping, tissue typing, preventive medicine, and pharmacogenomics. The

XX polypeptides are also useful as vaccines. This sequence represents an

XX example of the polypeptide of the invention.

SQ Sequence 2281 AA;

Query Match 93.4%; Score 11191; DB 8; Length 2281;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

Qy 1 MDPLIIFLLFIQTSETQVDVSNVVGTRDYDTISSISTYTSPTVTRIVPNVTKGPPV 60

Db 1 MDPLIIFLLFIQTSETQVDVSNVVGTRDYDTISSISTYTSPTVTRIGASN--EGGPPV 58

Qy 61 FLAGERVGSAGILLSNWTPPNNGRIISIVIKYKEVCPMWTYITQVRKPSLEVLTN 120

Db 59 FLAGERVGSAGILLSNWTPPNNGRIISIVIKYKEVCPMWTYITQVRKPSLEVLTN 118

QY 121 LNPCTTYEIKVAARENSAGIVPSDPFLFQTAESAPGKVNLVTEAYNAGVAKLWTLPRQ 180
DB 119 LNPCTTYEIKVAARENSAGIVPSDPFLFQTAESAPGKVNDFTGEAVPSS-KLMWYTSAT 177
QY 181 PNGKITSFKISVKAHRSIGVWVDVSRVEDILTKLP-ECNENSESFLMSTASPSPTLGR 239
DB 178 KK-KITSFKISVKAHRSIGVWVDVSRVEDILTKLP-ECNENSESFLMSTASPSPTLGR 236
QY 240 VTPPSRTTHSSSTLTQNEISSVWKEPISFVVTTHLRLPYTYTLFVSAATTEAGYIDSTIVR 299
DB 237 VTPPSRTTHSSSTLTQNEISSV-KEPISFVVTTHLRLPYTYTLFVSAATTEAGYIDSTIVR 295
QY 300 TPESVPEGPONCVTGNITGKSFILWDPTIITGKFSYRVELYGPS-GRILDNSTKDLK 358
DB 296 TPESVPEGPONCVTGNITGKSFILWDPTIITGKFSYRVELYGPSAGRILDNSTKDLK 355
QY 359 FAFTNLTPTFMYDVYIAAETSAGTGPKSNISVFTPPDVGAVPDLQAEVSTQVRITWK 418
DB 356 FAFTNLTPTFMYDVYIAAETSAGTGPKSNISVFTPPDVGAVPDLQAEVSTQVRITWK 415
QY 419 KRPQNGIINOVRKVLVPETGIIILENTLLTGNEYINDMPAPEIYNIIVPMVGLYEGSA 478
DB 416 KRPQNGIINOVRKVLVPETGIIILENTLLTGNEY-INDMPAPEIYNIIVPMVGLYEGSA 474
QY 479 EMSDDLHSLATFIYNHSHPKNPPARNRAEDQTSVVTTRNOYITDIAAQLSVIIRLVP 538
DB 475 EMSDDLHSLATFIYNHSHPKNPPARNRAEDQTSVVTTRNOYITDIAAQLSVIIRLRR 534
QY 539 FTEHMLSVSAFTIMSGPPVNLVSRTRQOVPSIKIINX--NISSSILLWDPPEYN 596
DB 535 FWAETHGFRYTIMSS-----ASRDNLTSPGLSAQNFVRVTHVITEVFLHWDPPD--P 586
QY 597 GKITHYTIWAMELDT-NRAFOITIDN-SFLITGLKKYTKYKMRVAASVTHGESSISEEN 654
DB 587 VFFHHVLIITLDVENQSKSILRTLSLSVLGLKKYTKYKMRVAASVTHGESSISEEN 646
QY 655 DIFVRTSEBPESSPODVEIDVTADEIRLKWSPPEKPNGLIIAYEVLKNDITLYMKN 714
DB 647 DIFVRTSEBPESSPODVEIDVTADEIRLKWSPPEKPNGLIIAYEVLKNDITLYMKN 706
QY 715 STTDIILRLNRLHLYNLSVRSVTRFCHGNOVSSLLSVTSETVPSAPENITYKNISSG 774
DB 707 STTDIILRLNRLHLYNLSVRSVTRFCHGNOVSSLLSVTSETVPSAPENITYKNISSG 766
QY 775 EIELSFLPSSPNGIIXKTYIYLRSGNEERTINTSLTQNIKVLKKTQVYIIEVSAST 834
DB 767 EIELSFLPSSPNGIIXKTYIYLRSGNEERTINTSLTQNIKVLKKTQVYIIEVSAST 826
QY 835 LKGEVRSAPIISILTEEDAPSPQDFSVKQLSGVTVKLSWQPLEPNGIILYTYVYWN 894
DB 827 LKGEVRSAPIISILTEEDAPSPQDFSVKQLSGVTVKLSWQPLEPNGIILYTYVYWN- 885
QY 895 RSSLKTINTETLSLESLDLYNVEYSAYTASTRFGDKTGNSIISFQTPGAPSPDKD 954
DB 886 RSSLKTINTETLSLESLDLYNVEYSAYTASTRFGDKTGNSIISFQTPGAPSPDKD 944
QY 955 VYVYANLSSSIIILFWTPPSKPNGLIOYYSVYVYRNTSGTFMONTLHENDFDMTVSTI 1014
DB 945 VYVYANLSSSIIILFWTPPSKPNGLIOYYSVYVYRNTSGTFMONTLHENDFDMTVSTI 1004
QY 1015 IDKLTIFFSYTFLWTASTSVGNKSSDIIEVYTDQDIPGFGVGNLTYESISSTAINVSW 1074
DB 1005 IDKLTIFFSYTFLWTASTSVGNKSSDIIEVYTDQDIPGFGVGNLTYESISSTAINVSW 1064
QY 1075 VPPAQPNGLVFTVYVLSILOQTPRHVRPPLVTVYRSIYFONLEKYTDYILKITPSTKGPS 1134
DB 1065 VPPAQPNGLVFTVYVLSILOQTPRHVRPPLVTVYRSIYFONLEKYTDYILKITPSTKGPS 1124
QY 1135 DRYTAQLYIKTBEDVPETSPINTFNKLSSTVLSWDPVPKPNCAIISYDITLQGNEN 1194
DB 1125 DRYTAQLYIKTBEDVPETSPINTFNKLSSTVLSWDPVPKPNCAIISYDITLQGNEN 1184
QY 1195 YSFITSDNYIILBELSPFLYFFFAAARTRKGLGPSSILFFYTDSESVPLAPPQNLTLINC 1254

DB 1185 YSFITSDNYIILBELSPFLYFFFAAARTRKGLGPSSILFFYTDSESVPLAPPQNLTLINC 1244
QY 1255 TSDFWLWKSPLPGGIVKVYSFKIHEHETDIYKNIISGPKTEAKLVGLEPVSYSIR 1314
DB 1245 TSDFWLWKSPLPGGIVKVYSFKIHEHETDIYKNIISGPKTEAKLVGLEPVSYSIR 1304
QY 1315 VSAFTKVGNGNQSNNVVKFTTQESVDPVQVONQCMATSWQSVLVKWDPPKANGIITQYM 1374
DB 1305 VSAFTKVGNGNQSNNVVKFTTQESVDPVQVONQCMATSWQSVLVKWDPPKANGIITQYM 1364
QY 1375 VTVERNSTKVSPODHYMTPIKLLANTSYVFKVRASTAGSESTCHVSTLPETVSVPT 1434
DB 1365 VTVERNSTKVSPODHYMTPIKLLANTSYVFKVRASTAGSESTCHVSTLPETVSVPT 1424
QY 1435 NIAFSDVOSTSATLWIRPDTILGYFQNYKIITQLRAACKWESECEVEYQIKOYLVEA 1494
DB 1425 NIAFSDVOSTSATLWIRPDTILGYFQNYKIITQLRAACKWESECEVEYQIKOYLVEA 1484
QY 1495 HLTETVYGLKKFRWYRFOVAASTNAGYGNASNIISTKTLPDPPDPPENHVAVTSPPS 1554
DB 1485 HLTETVYGLKKFRWYRFOVAASTNAGYGNASNIISTKTLPDPPDPPENHVAVTSPPS 1544
QY 1555 ISISSEPAVITGPTCYLIDVKSVDNDEENISFIKSNENKTETIKOLEIFRYSVVIITA 1614
DB 1545 ISISSEPAVITGPTCYLIDVKSVDNDEENISFIKSNENKTETIKOLEIFRYSVVIITA 1604
QY 1615 FTGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTQFQLTFPLPSPQNGNI 1674
DB 1605 FTGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTQFQLTFPLPSPQNGNI 1664
QY 1675 QYQALVYREDDPTAVQIHNLSIIQKNTFVJAMLEGLKGGHTYNIISVYAVNSAGAPKV 1734
DB 1665 QYQALVYREDDPTAVQIHNLSIIQKNTFVJAMLEGLKGGHTYNIISVYAVNSAGAPKV 1724
QY 1735 PNRITWIDKAPARPKTPTPIYDATGKLLVTSSTTITRMPICIYSDDHGPIKNVQVLATE 1794
DB 1725 PNRITWIDKAPARPKTPTPIYDATGKLLVTSSTTITRMPICIYSDDHGPIKNVQVLATE 1784
QY 1795 TGAQHDGNTWKYDAYFNKARPYFTNEGPNPCTEGKTFSGNEBIYIIGADNACWIPG 1854
DB 1785 TGAQHDGNTWKYDAYFNKARPYFTNEGPNPCTEGKTFSGNEBIYIIGADNACWIPG 1844
QY 1855 NEDKICNGPLKPKQYLFKFRATNIMGQFSDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1914
DB 1845 NEDKICNGPLKPKQYLFKFRATNIMGQFSDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1904
QY 1915 LSIILLGTAIFAFARIROKQEGGYSPQDABIIDTKLKLQDIIITVADILEKDERLTRPI 1974
DB 1905 LSIILLGTAIFAFARIROKQEGGYSPQDABIIDTKLKLQDIIITVADILEKDERLTRPI 1964
QY 1975 SKKSFLOHVEELCTNNLKFQEEFSELPKFLODLSSTADLPNNRAKRPFIKPNNNN 2034
DB 1965 SKKSFLOHVEELCTNNLKFQEEFSELPKFLODLSSTADLPNNRAKRPFIKPNNNN 2023
QY 2035 RVKLADASVPGSDYINASYISGYLCPNEFIATQCPGPGTGVDFWRVWVETRAKTLVMLT 2094
DB 2024 RVKLADASVPGSDYINASYISGYLCPNEFIATQCPGPGTGVDFWRVWVETRAKTLVMLT 2083
QY 2095 QCFEKGRIRCHQYWPEDNKPVTVFGDIIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCN 2154
DB 2084 QCFEKGRIRCHQYWPEDNKPVTVFGDIIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCN 2143
QY 2155 FTAWPEHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLQIND 2214
DB 2144 FTAWPEHGVSPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLQIND 2203
QY 2215 HDPVDIYGLVIELRSECMQNLQAOYIFLHCIIIDLLSNKGSNOPICFVNTSALQKQMS 2274
DB 2204 HDPVDIYGLVIELRSECMQNLQAOYIFLHCIIIDLLSNKGSNOPICFVNTSALQKQMS 2263
QY 2275 LDAME-GOVELEWSETTM 2291
DB 2263 LDAME-GOVELEWSETTM 2291

Db 2264 LDAMEGGDVELEWEETM 2281

RESULT 9

ID ADH41613

ADH41613 standard; protein; 2281 AA.

AC ADH41613;

XX DT 25-MAR-2004 (first entry)

XX DE Novel human protein NOV15a.

XX KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;

XX KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;

XX KW antiparkinsonian; antiasthmatic; antinfertility; cardiomyopathy;

XX KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;

XX KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;

XX KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;

XX KW tissue typing; preventive medicine; pharmacogenomic; vaccine.

OS Homo sapiens.

XX WO2003102159-A2.

XX 11-DEC-2003.

XX 04-JUN-2003; 2003WO-US017573.

XX 04-JUN-2002; 2002US-0385490P.

XX 04-JUN-2002; 2002US-0385615P.

XX 04-JUN-2002; 2002US-0385755P.

XX 05-JUN-2002; 2002US-0386041P.

XX 06-JUN-2002; 2002US-0386355P.

XX 06-JUN-2002; 2002US-0386357P.

XX 06-JUN-2002; 2002US-0386447P.

XX 06-JUN-2002; 2002US-0386459P.

XX 06-JUN-2002; 2002US-0386465P.

XX 06-JUN-2002; 2002US-0386864P.

XX 07-JUN-2002; 2002US-0386701P.

XX 07-JUN-2002; 2002US-0386796P.

XX 07-JUN-2002; 2002US-0386931P.

XX 07-JUN-2002; 2002US-0387078P.

XX 07-JUN-2002; 2002US-0387081P.

XX 07-JUN-2002; 2002US-0387083P.

XX 10-JUN-2002; 2002US-0387429P.

XX 10-JUN-2002; 2002US-0387540P.

XX 10-JUN-2002; 2002US-0387866P.

XX 11-JUN-2002; 2002US-0387606P.

XX 11-JUN-2002; 2002US-0387610P.

XX 11-JUN-2002; 2002US-0387659P.

XX 11-JUN-2002; 2002US-0387668P.

XX 11-JUN-2002; 2002US-0387696P.

XX 11-JUN-2002; 2002US-0387859P.

XX 12-JUN-2002; 2002US-0387934P.

XX 12-JUN-2002; 2002US-0387960P.

XX 12-JUN-2002; 2002US-0388022P.

XX 12-JUN-2002; 2002US-0388096P.

XX 12-JUN-2002; 2002US-0388432P.

XX 12-JUN-2002; 2002US-0388479P.

XX 13-JUN-2002; 2002US-0389123P.

XX 14-JUN-2002; 2002US-0389120P.

XX 14-JUN-2002; 2002US-0389146P.

XX 17-JUN-2002; 2002US-0389742P.

XX 18-JUN-2002; 2002US-0389604P.

XX 18-JUN-2002; 2002US-0389884P.

XX 19-JUN-2002; 2002US-0390006P.

XX 19-JUN-2002; 2002US-0390144P.

XX 19-JUN-2002; 2002US-0390209P.

XX 25-JUN-2002; 2002US-0391726P.

XX 06-AUG-2002; 2002US-0401628P.

XX 09-AUG-2002; 2002US-0402268P.

XX 12-AUG-2002; 2002US-0402822P.

PR 13-AUG-2002; 2002US-0403458P.

PR 15-AUG-2002; 2002US-0403617P.

PR 15-AUG-2002; 2002US-0403732P.

PR 26-AUG-2002; 2002US-0406182P.

PR 12-SEP-2002; 2002US-0410085P.

PR 13-SEP-2002; 2002US-0410505P.

PR 23-SEP-2002; 2002US-0412955P.

PR 30-SEP-2002; 2002US-0415195P.

PR 23-OCT-2002; 2002US-0420627P.

PR 23-OCT-2002; 2002US-0420718P.

PR 24-OCT-2002; 2002US-0420852P.

PR 31-OCT-2002; 2002US-0422750P.

PR 01-NOV-2002; 2002US-0423095P.

PR 05-NOV-2002; 2002US-0423748P.

XX (CURA-) CURAGEN CORP.

XX Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;

PI Burgess CE, Casman SJ, Catterton E, Dhanabai M, Edinger SR;

PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;

PI Grose WM, Gunther E, Guo X, Gusev VY, Hermann JL, Ji W, Kekuda R;

PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;

PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;

PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;

PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;

PI Wolenc AR, Zhong M, Zhong H;

XX WPI; 2004-053467/05.

DR N-PSDB; ADH41612.

XX New NOVX polypeptides and nucleic acid molecules useful for preventing or

PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,

PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

PT pharmacogenomics.

XX Claim 2; SEQ ID NO 166; 1503pp; English.

XX The invention relates to 566 new isolated human polypeptides and their

CC encoding genes, sequences that are at least 95% identical to these or

CC sequences comprising one or more conservative substitutions in these. The

CC polypeptide, polynucleotide and antibodies against the polypeptides are

CC useful in diagnosing, treating or preventing NOVX-associated disorders,

CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,

CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,

CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.

CC The nucleic acids are further used as hybridization probes, in chromosome

CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The

CC polypeptides are also useful as vaccines. This sequence represents an

CC example of the polypeptide of the invention.

XX Sequence 2281 AA;

SQ

Query Match 93.4%; Score 11191; DB 8; Length 2281;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

QY 1 MDFLIFLLFFLFGTSETQVDVSNVVPCTRYDITISSITTYTSPVTRIVTPNVTKEGPPV 60

DB 1 MDFLIFLLFFLFGTSETQVDVSNVVPCTRYDITISSITTYTSPVTRIVTPNVTKEGPPV 58

QY 61 FLAHERVGSAGILLSWNTPPNPNGRISIVYKVKVCPMMQVTVYTVQVRKSPDSLEVLITN 120

DB 59 FLAHERVGSAGILLSWNTPPNPNGRISIVYKVKVCPMMQVTVYTVQVRKSPDSLEVLITN 118

QY 121 LNPGETTYEIKVAENSAGIGVFSDFPLFQTASSAPGKVNLTVEAYNASAVKLIWYLPQ 180

DB 119 LNPGETTYEIKVAENSAGIGVFSDFPLFQTASSAPGKVNLTVEAYNASAVKLIWYLPQ 177

QY 181 PNGKITSPKISVKHARSIGIVKQVSRVEDILTGLP-ECNENSESEFLMSTASPSPTLGR 239

DB 178 KK-KITSPKISVKHRSIGIVKQVSRVEDILTGLP-ECNENSESEFLMSTASPSPTLGR 236

QY 240 VTPSRTHSSSTLTONISSVWKEPISPVWTHLRPYTYTLPEVSAATTEAGYIDSTIVR 299
DB 237 VTPSRTHSSSTLTONISSV--KEPISFVWTHLRPYTYTLPEVSAATTEAGYIDSTIVR 295
QY 300 TPESVPEGPPONCVGNIGTKSPSILWDPTIIVTGKFSVRVELYGPS-GRILDNSTKOLK 358
DB 296 TPESVPEGPPONCVGNIGTKSPSILWDPTIIVTGKFSVRVELYGPSAGRIILDNSTKOLK 355
QY 359 FAPTNTPTPMYDVYIAAETSAGTGPKNISVFTPPDPGAVPDLQALAEVESTQVRIITWK 418
DB 356 FAPTNTPTPMYDVYIAAETSAGTGPKNISVFTPPDPGAVPDLQALAEVESTQVRIITWK 415
QY 419 KRPQNGIINOYRVKVLVPETGIIILENTLLTGNEVINDPMAPEIYNIPEWGLYEGSA 478
DB 416 KRPQNGIINOYRVKVLVPETGIIILENTLLTGNE--INDPMAPEIYNIPEWGLYEGSA 474
QY 479 EMSDDLHSLATFIYNHSHDKNPPARNRAEDQTSVVTTRNOYITDIAAQLSIVIRLVP 538
DB 475 EMSDDLHSLATFIYNHSHDKNPPARNRAEDQTSVVTTRNOYITDIAAQLSIVIRLVP 534
QY 539 FTEHMISSVAPTIMGEGPPPTVLSVRTRQQVPSSIKIINYK--NISSSSILLWDPPEYPN 596
DB 535 FWAETWGFGRYTIMSS-----ASRONLTSPGLSAQNFRVTHVITTEVFLHWDPPD--P 586
QY 597 GKITHYTIYAMELDT-NRAPOITIDN-SFLITGLKKYTKYKRVAASTHDESSISEEN 654
DB 587 VPFHHYLLITLDVENQSKSIILATLMSLSLVGLKKYTKYKRVAASTHDESSISEEN 646
QY 655 DIFVRTSEBEPSSPOQVEIDVTADIEIRLKWSPPEKPNIGIIAYEVLYKNIDTLWKN 714
DB 647 DIFVRTSEBEPSSPOQVEIDVTADIEIRLKWSPPEKPNIGIIAYEVLYKNIDTLWKN 706
QY 715 STTDIILRLNRLPHTLYNISVRSYTRFGHGNQVSSLSVRTSEVPDSAPENITYKNISSG 774
DB 707 STTDIILRLNRLPHTLYNISVRSYTRFGHGNQVSSLSVRTSEVPDSAPENITYKNISSG 766
QY 775 EIELSLFPSSPNGIILKYYIILKRSNGHEERTINTSLTONIKVLKKTQYIIEVSAST 834
DB 767 EIELSLFPSSPNGIILKYYIILKRSNGHEERTINTSLTONIKVLKKTQYIIEVSAST 826
QY 835 LKGEGRSAPISILITBEDAPDSPQDPFSVKQLSGVTVKLSWQPLSPNGIILYYTYVKN 894
DB 827 LKGEGRSAPISILITBEDAPDSPQDPFSVKQLSGVTVKLSWQPLSPNGIILYYTYVKN 885
QY 895 RSSLKTINTVETLSLEISLDLYNVEYSAYVTASTFRFGDGKTSNIIISFQTEGAPSDPPKD 954
DB 886 RSSLKTINTVETLSLEISLDLYNVEYSAYVTASTFRFGDGKTSNIIISFQTEGAPSDPPKD 944
QY 955 VYVANISSSIIILFWTPPSKPNIGIIQYISVYVYNTSGTFMONTLHENDPDMVTSTI 1014
DB 945 VYVANISSSIIILFWTPPSKPNIGIIQYISVYVYNTSGTFMONTLHENDPDMVTSTI 1004
QY 1015 IDKLTIFSYTYFWLTASTSVGNKSSDILEVYTDQDPEGVGNLTYSISSTAINVSM 1074
DB 1005 IDKLTIFSYTYFWLTASTSVGNKSSDILEVYTDQDPEGVGNLTYSISSTAINVSM 1064
QY 1075 VPPAQPNGLVYVYVLSILOQTPHVRPPLVYTERSIFDNLEKYTDYILKITPSTTEKGPS 1134
DB 1065 VPPAQPNGLVYVYVLSILOQTPHVRPPLVYTERSIFDNLEKYTDYILKITPSTTEKGPS 1124
QY 1135 DYTQAQYIKTEEDVPETSPINTFNKLSGSTSVLLSWDPPVPKNGAIISYDLTLOQPNEN 1194
DB 1125 DYTQAQYIKTEEDVPETSPINTFNKLSGSTSVLLSWDPPVPKNGAIISYDLTLOQPNEN 1184
QY 1195 YSFIITSDNYIILELSPFTLYSFAAARTKGLGPSIILPFYTDSEVPLAPPONLTILNC 1254
DB 1185 YSFIITSDNYIILELSPFTLYSFAAARTKGLGPSIILPFYTDSEVPLAPPONLTILNC 1244
QY 1255 TSDPFWLKWSPSPPLPGIIVKVSFKIHEHETDIYIYKNISGFKTEAKLVGLEBPVSTYSIR 1314
DB 1245 TSDPFWLKWSPSPPLPGIIVKVSFKIHEHETDIYIYKNISGFKTEAKLVGLEBPVSTYSIR 1304
QY 1315 VSAFTKVGNGNQSNNVVKFTTQSSVDPVQNMCMATSWQSVLVKWDPPKANGIITQYM 1374

RESULT 10
ABP60058
ID ABP60058 standard; protein; 2301 AA.
XX
AC ABP60058;

DB 1305 VSAFTKVGNGNQSNNVVKFTTQSSVDPVQNMCMATSWQSVLVKWDPPKANGIITQYM 1364
QY 1375 VTVERNSTKVSQDGHMYTFIKLLANTSYYVKVRASTASAGEDESTCHVSTLTPETVPSVPT 1434
DB 1365 VTVERNSTKVSQDGHMYTFIKLLANTSYYVKVRASTASAGEDESTCHVSTLTPETVPSVPT 1424
QY 1435 NIAFSDVQSTSATLWIRPDTTILGFQNYKITTLQRAQCKEWESECEYQIKIQLYLEA 1494
DB 1425 NIAFSDVQSTSATLWIRPDTTILGFQNYKITTLQRAQCKEWESECEYQIKIQLYLEA 1484
QY 1495 HLTETVYGLKKFRVYRFOVAASTNAGYGNASNWISTKTLPDPGPPNHHVAVTSPS 1554
DB 1485 HLTETVYGLKKFRVYRFOVAASTNAGYGNASNWISTKTLPDPGPPNHHVAVTSPS 1544
QY 1555 ISISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIIEIKDLBIFRYSVVI 1614
DB 1545 ISISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIIEIKDLBIFRYSVVI 1604
QY 1615 FTGNIISAAVVEGKSSAEMI VTTLESAPKDPNNMTFQKI PDEVTKFQLTFLPPSQPNGI 1674
DB 1605 FTGNIISAAVVEGKSSAEMI VTTLESAPKDPNNMTFQKI PDEVTKFQLTFLPPSQPNGI 1664
QY 1675 QYQALVYREDDPTAVQIHNLISIIOKTNTFVIAMLEGLKGGHTYNI SVAVANSAGAGPKV 1734
DB 1665 QYQALVYREDDPTAVQIHNLISIIOKTNTFVIAMLEGLKGGHTYNI SVAVANSAGAGPKV 1724
QY 1735 PMRIITMDIKAPARPKTPTPIYDATGKLLAVTSTTITIRMPICYSDHGPINKNVVLATE 1794
DB 1725 PMRIITMDIKAPARPKTPTPIYDATGKLLAVTSTTITIRMPICYSDHGPINKNVVLATE 1784
QY 1795 TGAQHDGNVTKWYDAYFNKARPYFTNEGPPNPCTEGTKFSGNEBIYIIGADNACMIPG 1854
DB 1785 TGAQHDGNVTKWYDAYFNKARPYFTNEGPPNPCTEGTKFSGNEBIYIIGADNACMIPG 1844
QY 1855 NEDKICNGPLKPKQYLFKFRATNIMGQPTSDSYSDPVKTLGEGLSERTVEIILSVTLCI 1914
DB 1845 NEDKICNGPLKPKQYLFKFRATNIMGQPTSDSYSDPVKTLGEGLSERTVEIILSVTLCI 1904
QY 1915 LSIILLGTAFAPARIROKQEGGTVSPQDAI IITDKLDQIITVADLELKDRLTRPI 1974
DB 1905 LSIILLGTAFAPARIROKQEGGTVSPQDAI IITDKLDQIITVADLELKDRLTRPI 1964
QY 1975 SKKSFLOHVEELCTNNLKFQBEFSEL PKFLODLSSTADLPMNRAKNRFPNIPYNNNN 2034
DB 1965 SKKSFLOHVEELCTNNLKFQBEFSEL PKFLODLSSTADLPMNRAKNRFPNIPYNNNN 2023
QY 2035 RVKLTADASVPGSDYINASYISGYLCPNEFIATQGPLCTVGDVFWMMWETRAKTLVMLT 2094
DB 2024 RVKLTADASVPGSDYINASYISGYLCPNEFIATQGPLCTVGDVFWMMWETRAKTLVMLT 2083
QY 2095 QCFEKGRIRCHOYWPEDNKPVTVFGDI VITKLMEDVQIDWTIRDLKIERHGDGMTVRQCN 2154
DB 2084 QCFEKGRIRCHOYWPEDNKPVTVFGDI VITKLMEDVQIDWTIRDLKIERHGDGMTVRQCN 2143
QY 2155 FTAWPEHGVPENSAPLIHFVKLVRASTADHTTTPMIVHCSAGVGTGVTFIADHLTQHIND 2214
DB 2144 FTAWPEHGVPENSAPLIHFVKLVRASTADHTTTPMIVHCSAGVGTGVTFIADHLTQHIND 2203
QY 2215 HDPVDIYGLVABLRSERMCMQNLQAYIFLHCILDLNLSKNSQNPICFVNTYSALQKMS 2274
DB 2204 HDPVDIYGLVABLRSERMCMQNLQAYIFLHCILDLNLSKNSQNPICFVNTYSALQKMS 2263
QY 2275 LDAME-GDVELEWEETTM 2291
DB 2264 LDAMEGDVELEWEETTM 2281


```
QY 1379 RNSTKVSPODHMTFKILLANTSYVFKVASTASAGBDRSTCHVSTLPTVPSVPTNIAP 1438
DB 1381 GNSKTVSPRDPYTFKLLPNTSYVFEVRASTASAGBESRCDISTLPTVPSAPTNAV 1440
QY 1439 SDVQSTASATLWIRPDTILGYFQNYKITTLQRAQCKWESECEVEYQKIQYLYEAHLE 1498
DB 1441 SNVQSTASATLWIRPDTILGYFQNYKITTLQRAQCKWESECEVEYQKIQYLYEAHLE 1500
QY 1499 ETYVGLKPRWRFRQVAASTNAGYGNASNMI STKLPGLPPDPENHVHVAATSPFSIS 1558
DB 1501 ETVHGLKPRWRFRQVAASTNAGYGNASNMI STQTLPGPPDPENHVHVAATSPFGINS 1560
QY 1559 WSEPAVITGPTCVLIDVKSVDNDSFNI SPFKNEENKTTIEIKDLLEIFTRYSVITAFGN 1618
DB 1561 WSEPAVITGPTCVLIDVKSVDNDSFNI SPFKNEENKTTIEINNLEVFTRYSVITAFGN 1620
QY 1619 ISAAVYEGKSSAEIVITLESAPKPPNMTFKIPDRVTKFQLTFLPPSQPNQNIQVQ 1678
DB 1621 VSRAYTDGKSSAEIVITLESVPKPPNMTFKIPDRVTKFQLTFLPPSQPNQNIQVQ 1680
QY 1679 ALVYREDDPTAVQIHNLSIIQKNTFVIAMLEGLKGHTYNI SVAVNSAGAGPKVPMRI 1738
DB 1681 ALVYREDDPTAVQIHNLSIIQKNTFVIAMLEGLKGHTYNI SVAVNSAGAGPKVPMRI 1740
QY 1739 TMDIKAPARKPTPIYDATGKLLVSTTITIRMPICYSDHGGPIKRVQVLATETGAQ 1798
DB 1741 TMDIKAPARKPTPIYDATGKLLVSTTITIRMPICYSDHGGPIKRVQVLATETGAQ 1800
QY 1799 HGNVTKWDAVFNKARPYFTNEGPNPCTEGKTFKESNEEYIIGADNACWIPGNEBK 1858
DB 1801 QDGNVTKWDAVFNKARPYFTNEGPNPCTEGKTFKESNEEYIIGADNACWIPGNEBK 1860
QY 1859 ICNGPLKPKQVLFKFRATNIMGQTDSDYSPVKTGLGSLERTVEIILSVTLCLISII 1918
DB 1861 ICNGPLKPKQVLFKFRATNIMGQTDSDYSPVKTGLGSLERTVEIILSVTLCLISII 1920
QY 1919 LIGTAIFAPARIQKQEGGTSYSPQDAEIIDTKLKLQDITVADLEKDERLTR----- 1972
DB 1921 LIGTAIFAPARIQKQEGGTSYSPQDAEIIDTKLKLQDITVADLEKDERLTR----- 1980
QY 1973 ---PISKSKFLOHVEELCTNNIKQEESESELPKFLDLSSTADLPWNRKRNPNIKP 2029
DB 1981 SIKPISKSKFLOHVEELCTNNIKQEESESELPKFLDLSSTADLPWNRKRNPNIKP 2040
QY 2030 YNNNRVKLIADASVPGSDYINASYISGVLCPNEPIATQGPLGTGDFRVMWETRAKT 2089
DB 2041 YNNNRVKLIADASVPGSDYINASYISGVLCPNEPIATQGPLGTGDFRVMWETRAKT 2099
QY 2090 LVMLTQCPEKGRIRCHQYWPEDNKPVTVFQDVIITKLMEDVQIDMTIRDLKIERHGDCT 2149
DB 2100 LVMLTQCPEKGRIRCHQYWPEDNKPVTVFQDVIITKLMEDVQIDMTIRDLKIERHGDCT 2159
QY 2150 VQOCNFTAMPBGVPENSAPLHFVKLVASRAHDTTMI VHCAGVGRGTGFIALDHLT 2209
DB 2160 VQOCNFTAMPBGVPENSAPLHFVKLVASRAHDTTMI VHCAGVGRGTGFIALDHLT 2219
QY 2210 QHINHDHFDYIYGLVAELSERCMVQNLQAOYIFLHQCTILDLSNKGNOPICFVNYSL 2269
DB 2220 QHINHDHFDYIYGLVAELSERCMVQNLQAOYIFLHQCTILDLSNKGNOPICFVNYSL 2279
QY 2270 QKMDSLDAMEGDVLEWEETTM 2291
DB 2280 QKMDSLDAMEGDVLEWEETTM 2301
```

RESULT 11

AAE37322

ID AAE37322 standard; protein; 1959 AA.

XX AAE37322;

AC AAE37322;

XX 07-AUG-2003 (first entry)

```
XX Human receptor tyrosine kinase protein #3.
DE Human; receptor tyrosine phosphatase; diabetes; obesity; CNS disorder;
XX multiple sclerosis; epilepsy; chronic obstructive pulmonary disease;
KW COPD; Parkinson's disease; stroke; cardiovascular disorder; arrhythmia;
KW congestive heart failure; myocardial infarction; ischaemic heart disease;
KW gene therapy; anorectic; cardiant; neuroprotective; anticonvulsant;
KW cerebroprotective; vasotropic; antiarrhythmic; receptor; enzyme.
XX Homo sapiens.
OS WO2003033688-A1.
XX 24-APR-2003.
XX 14-OCT-2002; 2002WO-EP011473.
XX 16-OCT-2001; 2001US-0329329P.
XX (FARB ) BAYER AG.
XX Liou J;
XX WPI; 2003-403215/38.
XX N-PSDB; AAD56413.
XX Novel polynucleotides encoding human receptor tyrosine phosphatase
PT polypeptides, useful for treating diabetes, CNS disorders, obesity,
PT chronic obstructive pulmonary disease and cardiovascular disorders.
XX PS Disclosure; Fig 5; 163pp; English.
XX The present invention relates to receptor tyrosine phosphatase proteins
CC and polynucleotides encoding them. Sequences of the invention are useful
CC in the preparation of medicaments for modulating the activity of receptor
CC tyrosine phosphatase in disease such as diabetes, obesity, CNS disorders
CC (multiple sclerosis, epilepsy, Parkinson's disease and stroke), chronic
CC obstructive pulmonary disease (COPD) and cardiovascular disorders (e.g.
CC congestive heart failure, myocardial infarction, ischaemic heart disease
CC and arrhythmia). They are also used in gene therapy. The present sequence
CC is human receptor tyrosine kinase protein
XX Sequence 1959 AA;
SQ
Query Match 82.4%; Score 9874; DB 6; Length 1959;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 1998; Conservative 15; Mismatches 29; Indels 36; Gaps 4;
QY 331 IVTKFYSYRVLYGPSRI-----LDNSTKDLKFAFTNLTPFTMYDVYIAAETSAGT 382
DB 1 LLGGKLTNRKDIHTKNPSVHHHQRPKVDKTK-----MGKKQSRKT 42
QY 383 GPKSNISVFTPPDVGAVFDLQAEVSTQVRI TWKKRPQNGIINQYRVKLVLPETGII 442
DB 43 GNSKQSTSPPPKQGRAVFDLQAEVSTQVRI TWKKRPQNGIINQYRVKLVLPETGII 102
QY 443 LENTLLTGNEYNDMPAPEI VIVPEPMVGLYEGSAEMSSDLHSLATFIYNHPDKNFFA 502
DB 103 LENTLLTGNEYNDMPAPEI VIVPEPMVGLYEGSAEMSSDLHSLATFIYNHPDKNFFA 162
QY 503 RNRAEDQTSPPVVTTRNQYITDIAEQLSYVIRRLVPFTTEHMI SVSAFTMGSPPTVLSV 562
DB 163 RNRAEDQTSPPVVTTRNQYITDIAEQLSYVIRRLVPFTTEHMI SVSAFTMGSPPTVLSV 222
QY 563 RTRQOVSPSSIKIINTKNISSSILLYWDPPEYNGKI THYTYIAMELDTNRAFOITTDIN 622
DB 223 RTRQOVSPSSIKIINTKNISSSILLYWDPPEYNGKI THYTYIAMELDTNRAFOITTDIN 282
QY 623 SPLITGLKXKTKYKMRVAASHTDGESSLSEENDI FVRTSEDEPSSPODVEVIDVTADRI 682
DB 283 SPLITGLKXKTKYKMRVAASHTDGESSLSEENDI FVRTSEDEPSSPODVEVIDVTADRI 342
```

Qy	683	RLKWSPEKPNGLIIAYEVLYKNIIDTILYMKNTSTTDIILRNLRPHITLYNLSVRSYTRFGH	742
Db	343	RLKWSPEKPNGLIIAYEVLYKNIIDTILYMKNTSTTDIILRNLRPHITLYNLSVRSYTRFGH	402
Qy	743	GNQVSSLLSVRTSETVPDSAPENITYKNIISGGETLSFLPSSPNGIIKKYTYLKRKSN	802
Db	403	GNQVSSLLSVRTSETVPDSAPENITYKNIISGGETLSFLPSSPNGIIKKYTYLKRKSN	462
Qy	803	NEERTINTSLTQNIKVLKKYTOYIIIEVSASTLKGEGRSAPISILTEEDAPSPPODFS	862
Db	463	NEERTINTSLTQNIKVLKKYTOYIIIEVSASTLKGEGRSAPISILTEEDAPSPPODFS	522
Qy	863	VKQLSGVTVKLSWQPPLEPGIILYYTYVWNRSSLKTINVTETSLBELSOLDYNNVEYSAY	922
Db	523	VKQLSGVTVKLSWQPPLEPGIILYYTYVWNRSSLKTINVTETSLBELSOLDYNNVEYSAY	582
Qy	923	VTASTRGDGTGKGNIIISFOTPEGAPSDPKDQVYANLSSSIIILFWTPSPKNGIIQY	982
Db	583	VTASTRGDGTGKGNIIISFOTPEGAPSDPKDQVYANLSSSIIILFWTPSPKNGIIQY	642
Qy	983	SVYYRNTSGTFMQNFTLHETNDNMVTSIIDKLTFSYTYTFWLASTSVGNKSSD	1042
Db	643	SVYYRNTSGTFMQNFTLHETNDNMVTSIIDKLTFSYTYTFWLASTSVGNKSSD	702
Qy	1043	IIIEVYTODDIPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFFYVLSIILOQTPHRVRPP	1102
Db	703	IIIEVYTODDIPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFFYVLSIILOQTPHRVRPP	762
Qy	1103	LVTYERSIYFDNLEKTYDYLKIITPSTKESDPTTAQLYIKTEEDVPETSPINTPKNL	1162
Db	763	LVTYERSIYFDNLEKTYDYLKIITPSTKESDPTTAQLYIKTEEDVPETSPINTPKNL	822
Qy	1163	SSTSVLLSWDPVPKPNGALISDULTLOQPNENYFITSNDVIIIEELSPFTLXSFPAAR	1222
Db	823	SSTSVLLSWDPVPKPNGALISDULTLOQPNENYFITSNDVIIIEELSPFTLXSFPAAR	882
Qy	1223	TRKGLGPSSILFFYTDESVP LAPQNLTLINCTSDFVWLKWSPSPLPGGIVKYVFXIHE	1282
Db	883	TRKGLGPSSILFFYTDESVP LAPQNLTLINCTSDFVWLKWSPSPLPGGIVKYVFXIHE	942
Qy	1283	HETDTIIYKNIISGFKTEAKLVGLEPVSTYSIRUSAPTKVGNQOFSNNVKFTTQESVPDV	1342
Db	943	HETDTIIYKNIISGFKTEAKLVGLEPVSTYSIRUSAPTKVGNQOFSNNVKFTTQESVPDV	1002
Qy	1343	VQNMQCHATSQCSVLVKWDPPKKANGIIITQMVTVERNSTKVSPODMYTFIKLLANTS	1402
Db	1003	VQNMQCHATSQCSVLVKWDPPKKANGIIITQMVTVERNSTKVSPODMYTFIKLLANTS	1062
Qy	1403	VFKVRASTSGEGDESTCHVSTLPETVPSVPTNIAFSDVOSTSATLTWIRPDTILGYFQN	1462
Db	1063	VFKVRASTSGEGDESTCHVSTLPETVPSVPTNIAFSDVOSTSATLTWIRPDTILGYFQN	1122
Qy	1463	YKIIITQIARAQCKEWESECEVEYOKI QYLYEABHITETVTYGLKFRWYRQVAASTWAG	1522
Db	1123	YKIIITQIARAQCKEWESECEVEYOKI QYLYEABHITETVTYGLKFRWYRQVAASTWAG	1182
Qy	1523	GNASNWISTKTLPGPPDGPNNVHVATSPESISISWSEPAVITGPTCYLIDVKSVDNDE	1582
Db	1183	GNASNWISTKTLPGPPDGPNNVHVATSPESISISWSEPAVITGPTCYLIDVKSVDNDE	1242
Qy	1583	FNISFIKSNENKTIIEIKDLIEIFRYSVITAFIGNISAAVVEGKSSAEMIVTTLESAPK	1642
Db	1243	FNISFIKSNENKTIIEIKDLIEIFRYSVITAFIGNISAAVVEGKSSAEMIVTTLESAPK	1302
Qy	1643	DPNNMTFQKIPDEVTKPQLFTFLPPSPQNGNIQYQALVYREDDPTAVQIHNLSIIOKTN	1702
Db	1303	DPNNMTFQKIPDEVTKPQLFTFLPPSPQNGNIQYQALVYREDDPTAVQIHNLSIIOKTN	1362
Qy	1703	TFVTAMLEGLKGGHTYILSVYAVNSAGPKVPMRIITMDIKAPARPKTKPTPIYDAGTKL	1762
Db	1363	TFVTAMLEGLKGGHTYILSVYAVNSAGPKVPMRIITMDIKAPARPKTKPTPIYDAGTKL	1422
Qy	1763	LVTSTTITIRMPICVYSDDHGPIKXNOVLATETGAQHDGNVTWKYDAYFNKARPYFTNEG	1822

Db	1423	LVSTTTITIRMPICYYSDDDGPPKKNQVVLVTETGAQHDGNTWKYDAIFNKARPYFTNEG	1482
Qy	1823	FPNPPTCEGKTGKSGNEEIIYIIGADNACMTPGNEDKICNGPLPKPKQYLFKFRATNIMGQ	1882
Db	1483	FPNPPTCEGKTGKSGNEEIIYIIGADNACMTPGNEDKICNGPLPKPKQYLFKFRATNIMGQ	1542
Qy	1883	FTDSYSDPVKTIIGEGLSERTVEIILSVTLCILSLILGTATAPAPARIROKQKEGGTYSYP	1942
Db	1543	FTDSYSDPVKTIIGEGLSERTVEIILSVTLCILSLILGTATAPAPARIROKQKEGGTYSYP	1602
Qy	1943	ODAEIIDTKLKDQLITVADLEKDBRLTR-----PISKKSFLQHVVEELCTNNLK	1993
Db	1603	ODAEIIDTKLKDQLITVADLEKDBRLTR-----PISKKSFLQHVVEELCTNNLK	1662
Qy	1994	FQEFSELPFLQDLSTADLPWNRANKRFPNPKPYNNNNRVKLIADASVPGSDYINAS	2053
Db	1663	FQEFSELPFLQDLSTADLPWNRANKRFPNPKPYNNNNRVKLIADASVPGSDYINAS	1721
Qy	2054	YISGYLCPNEFIATQGPLPGTVGDFWRMWETTRAKTLVMLTQCFEKGRIHQVWPEDNK	2113
Db	1722	YISGYLCPNEFIATQGPLPGTVGDFWRMWETTRAKTLVMLTQCFEKGRIHQVWPEDNK	1781
Qy	2114	PVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVROCNFTAWPBEHGVPENSAPLIHF	2173
Db	1782	PVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVROCNFTAWPBEHGVPENSAPLIHF	1841
Qy	2174	VKLVRASRAHDTTPMIVHCSAGVGRGTGFVFTALDHLTQHINDHDFVDIYGLVAELRSEMC	2233
Db	1842	VKLVRASRAHDTTPMIVHCSAGVGRGTGFVFTALDHLTQHINDHDFVDIYGLVAELRSEMC	1901
Qy	2234	MVQNLAQYIFLHQICILDLLSNKGSNPICFVNYSALQKMDSLDAMEGDVELEWEETM	2291
Db	1902	MVQNLAQYIFLHQICILDLLSNKGSNPICFVNYSALQKMDSLDAMEGDVELEWEETM	1959
RESULT 12			
ADH41621	ID	ADH41621 standard; protein; 909 AA.	
XX	AC	ADH41621;	
XX	DT	25-MAR-2004 (first entry)	
XX	DE	Novel human protein NOV15e.	
KW	KW	cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;	
KW	KW	antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;	
KW	KW	antiparkinsonian; antasthmatic; antifertility; cardiomyopathy;	
KW	KW	atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;	
KW	KW	multiple sclerosis; graft-versus-host disease; Alzheimer's disease;	
KW	KW	Parkinson's disease; asthma; fertility disorder; chromosome mapping;	
XX	XX	tissue typing; preventive medicine; pharmacogenomic; vaccine.	
OS	OS	Homo sapiens.	
XX	XX	WO2003102159-A2.	
XX	XX	11-DEC-2003.	
XX	XX	04-JUN-2003; 2003WO-US017573.	
XX	XX	04-JUN-2002; 2002US-0385490P.	
PR	PR	04-JUN-2002; 2002US-0385615P.	
PR	PR	04-JUN-2002; 2002US-0385755P.	
PR	PR	05-JUN-2002; 2002US-0386041P.	
PR	PR	06-JUN-2002; 2002US-0386355P.	
PR	PR	06-JUN-2002; 2002US-0386357P.	
PR	PR	06-JUN-2002; 2002US-0386447P.	
PR	PR	06-JUN-2002; 2002US-0386459P.	
PR	PR	06-JUN-2002; 2002US-0386465P.	
PR	PR	06-JUN-2002; 2002US-0386864P.	
PR	PR	07-JUN-2002; 2002US-0386701P.	

PR 07-JUN-2002; 2002US-0386796P.
 PR 07-JUN-2002; 2002US-0386931P.
 PR 07-JUN-2002; 2002US-0387078P.
 PR 07-JUN-2002; 2002US-0387081P.
 PR 07-JUN-2002; 2002US-0387083P.
 PR 10-JUN-2002; 2002US-0387429P.
 PR 10-JUN-2002; 2002US-0387540P.
 PR 10-JUN-2002; 2002US-0387866P.
 PR 11-JUN-2002; 2002US-0387606P.
 PR 11-JUN-2002; 2002US-0387610P.
 PR 11-JUN-2002; 2002US-0387659P.
 PR 11-JUN-2002; 2002US-0387688P.
 PR 11-JUN-2002; 2002US-0387896P.
 PR 11-JUN-2002; 2002US-0387899P.
 PR 12-JUN-2002; 2002US-0387934P.
 PR 12-JUN-2002; 2002US-0387960P.
 PR 12-JUN-2002; 2002US-0388022P.
 PR 12-JUN-2002; 2002US-0388096P.
 PR 12-JUN-2002; 2002US-0388432P.
 PR 13-JUN-2002; 2002US-0388479P.
 PR 14-JUN-2002; 2002US-0389123P.
 PR 14-JUN-2002; 2002US-0389146P.
 PR 17-JUN-2002; 2002US-0389742P.
 PR 18-JUN-2002; 2002US-0389604P.
 PR 18-JUN-2002; 2002US-0389884P.
 PR 19-JUN-2002; 2002US-0390006P.
 PR 19-JUN-2002; 2002US-0390144P.
 PR 19-JUN-2002; 2002US-0390209P.
 PR 25-JUN-2002; 2002US-0391726P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402268P.
 PR 12-AUG-2002; 2002US-0402822P.
 PR 13-AUG-2002; 2002US-0403458P.
 PR 15-AUG-2002; 2002US-0403617P.
 PR 15-AUG-2002; 2002US-0403732P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 12-SEP-2002; 2002US-0410085P.
 PR 13-SEP-2002; 2002US-0410505P.
 PR 23-SEP-2002; 2002US-0412955P.
 PR 30-SEP-2002; 2002US-0415195P.
 PR 23-OCT-2002; 2002US-0420627P.
 PR 23-OCT-2002; 2002US-0420718P.
 PR 24-OCT-2002; 2002US-0420852P.
 PR 31-OCT-2002; 2002US-0422750P.
 PR 01-NOV-2002; 2002US-0423095P.
 PR 05-NOV-2002; 2002US-0423748P.
 (CURA-) CURAGEN CORP.

Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL,
 Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR,
 Ellerman K, Ettenberg S, Gangoli EA, Gerlach VL, Gorman LJ,
 Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
 Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
 MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;
 Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
 Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G,
 Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
 Wolenc AR, Zhong M, Zhong H;

WPI: 2004-053467/05.
 N-PSDB; ADH41620.

New NOVX polypeptides and nucleic acid molecules useful for preventing or
 treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
 pharmacogenomics.

Claim 2; SEQ ID NO 174; 1503pp; English.

The invention relates to 566 new isolated human polypeptides and their
 encoding genes, sequences that are at least 95% identical to these or

CC sequences comprising one or more conservative substitutions in these. The
 CC polypeptide, polynucleotide and antibodies against the polypeptides are
 CC useful in diagnosing, treating or preventing NOVX-associated disorders,
 CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
 CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
 CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
 CC The nucleic acids are further used as hybridization probes, in chromosome
 CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
 CC polypeptides are also useful as vaccines. This sequence represents an
 CC example of the polypeptide of the invention.

XX Sequence 909 AA;

Query Match 38.9%; Score 4661; DB 8; Length 909;
 Best Local Similarity 99.6%; Pred. No. 1.9e-272;
 Matches 905; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 18 QVDVNVVPGTRDYDITISSITTYTPVTRIVTNVTKPGPVFLAGERVGSAGILLSMN 77
 DB 1 QVDVNVVPGTRDYDITISSITTYTPVTRIVTNVTKPGPVFLAGERVGSAGILLSMN 60
 QY 78 TTPNPNGRIIISIVKYKVCPMQVTVTVRSKPSLEVLTLNLPNGTYYEIKVAEENSA 137
 DB 61 TTPNPNGRIIISIVKYKVCPMQVTVTVRSKPSLEVLTLNLPNGTYYEIKVAEENSA 120
 QY 138 GIGVFSDDPFLQTAESAPGVNLTVEAYNASAVKLIWYLPQPNKGKITSFKISVKHARS 197
 DB 121 GIGVFSDDPFLQTAESAPGVNLTVEAYNASAVKLIWYLPQPNKGKITSFKISVKHARS 180
 QY 198 GIWVKDVSIRVEDILTGLKPECNENSESEFLWSTASPSPTLGRVTPSRTHSSSTLTQNE 257
 DB 181 GIWVKDVSIRVEDILTGLKPECNENSESEFLWSTASPSPTLGRVTPSRTHSSSTLTQNE 240
 QY 258 ISSVMKEPISFVTHLRPVTTTLFEVSAATTEAGYIDSTIVRTPESVPSPGPPQNCVTGNI 317
 DB 241 ISSVMKEPISFVTHLRPVTTTLFEVSAATTEAGYIDSTIVRTPESVPSPGPPQNCVTGNI 300
 QY 318 TGKSFSLWDPTTIVTGKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTWYDVYIAAE 377
 DB 301 TGKSFSLWDPTTIVTGKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTWYDVYIAAE 360
 QY 378 TSAGTGPKSNI SVFTPPDPVPGAVFDLQLAEVSTQVRIITWKKPQPNQINGIINQVRKVLVP 437
 DB 361 TSAGTGPKSNI SVFTPPDPVPGAVFDLQLAEVSTQVRIITWKKPQPNQINGIINQVRKVLVP 420
 QY 438 ETGIIENLTLLAGNNEYINDPNAPEIVNIVBPMVGLYEGSAEMSSDLHSLATFIYNHPD 497
 DB 421 ETGIIENLTLLAGNNEYINDPNAPEIVNIVBPMVGLYEGSAEMSSDLHSLATFIYNHPD 480
 QY 498 KNFPARNRAEDQTSVPVTTTRNQYITDIAAEQLSYVIRRLVPPFTEHNMISVSFAFTIMGEGPP 557
 DB 481 KNFPARNRAEDQTSVPVTTTRNQYITDIAAEQLSYVIRRLVPPFTEHNMISVSFAFTIMGEGPP 540
 QY 558 TVLSVTRTQOVPSKIKINYNKNISSSILLYWDPPYPNGKITHYTIYAMELDTNRAFI 617
 DB 541 TVLSVTRTQOVPSKIKINYNKNISSSILLYWDPPYPNGKITHYTIYAMELDTNRAFI 600
 QY 618 TTIDNSFLITGLKKYTKYKRVAASTHDCGESSISENDIFVRTSEDEPSSPODVVIDV 677
 DB 601 TTIDNSFLITGLKKYTKYKRVAASTHDCGESSISENDIFVRTSEDEPSSPODVVIDV 660
 QY 678 TADEIRLKWSPPEKNGIIIAVEVLYKNIDTLYMKNSTTIDILRLNRLPHTLYNISVRSY 737
 DB 661 TADEIRLKWSPPEKNGIIIAVEVLYKNIDTLYMKNSTTIDILRLNRLPHTLYNISVRSY 720
 QY 738 TRFGHGNQVSSLLSVRTSETVPSAPENITYKNISGEIELSFLPPSSPNGIKKTYIYL 797
 DB 721 TRFGHGNQVSSLLSVRTSETVPSAPENITYKNISGEIELSFLPPSSPNGIKKTYIYL 780
 QY 798 KRSNGNEERTINTTSLTQNIKVKYQYIIIEVSASTLKGEGVRSAPISILTEEDAPDSP 857
 DB 781 KRSNGNEERTINTTSLTQNIKVKYQYIIIEVSASTLKGEGVRSAPISILTEEDAPDSP 840

KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
 KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
 KW antiparkinsonian; antiaschmatic; antiinfertility; cardiomyopathy;
 KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
 KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
 KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
 KW tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX Homo sapiens.

XX WO2003102159-A2.

XX 11-DEC-2003.

XX 04-JUN-2003; 2003WO-US017573.

XX 04-JUN-2002; 2002US-0385490P.

XX 04-JUN-2002; 2002US-0385615P.

XX 04-JUN-2002; 2002US-0385755P.

XX 05-JUN-2002; 2002US-0386041P.

XX 06-JUN-2002; 2002US-0386355P.

XX 06-JUN-2002; 2002US-0386357P.

XX 06-JUN-2002; 2002US-0386447P.

XX 06-JUN-2002; 2002US-0386459P.

XX 06-JUN-2002; 2002US-0386465P.

XX 06-JUN-2002; 2002US-0386864P.

XX 07-JUN-2002; 2002US-0386701P.

XX 07-JUN-2002; 2002US-0386796P.

XX 07-JUN-2002; 2002US-0386931P.

XX 07-JUN-2002; 2002US-0387078P.

XX 07-JUN-2002; 2002US-0387081P.

XX 07-JUN-2002; 2002US-0387083P.

XX 10-JUN-2002; 2002US-0387429P.

XX 10-JUN-2002; 2002US-0387540P.

XX 10-JUN-2002; 2002US-0387866P.

XX 10-JUN-2002; 2002US-0387866P.

XX 11-JUN-2002; 2002US-0387806P.

XX 11-JUN-2002; 2002US-0387810P.

XX 11-JUN-2002; 2002US-0387659P.

XX 11-JUN-2002; 2002US-0387668P.

XX 11-JUN-2002; 2002US-0387696P.

XX 11-JUN-2002; 2002US-0387859P.

XX 12-JUN-2002; 2002US-0387934P.

XX 12-JUN-2002; 2002US-0387960P.

XX 12-JUN-2002; 2002US-0388022P.

XX 12-JUN-2002; 2002US-0388096P.

XX 12-JUN-2002; 2002US-0388432P.

XX 12-JUN-2002; 2002US-0388479P.

PR

05-NOV-2002; 2002US-0423748P.

XX

(CURA-) CURAGEN CORP.

PI

Alsbrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;

PI

Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;

PI

Ellerman K, Stettenberg S, Gangolli EA, Gerlach VL, Gorman L;

PI

Groshe WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;

PI

Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

PI

MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;

PI

Milliet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;

PI

Rieger DK, Rothenberg MB, Shenoy SG, Shimkets RA, Smithson G;

PI

Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;

PI

Wolenc AR, Zhong M, Zhong H;

XX

WPI; 2004-053467/05.

DR

N-PSDB; ADH41622.

XX

XX

PT

New NOVX polypeptides and nucleic acid molecules useful for preventing or

PT

treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,

PT

atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

PT

pharmacogenomics.

XX

XX

PS

Claim 2; SEQ ID NO 176; 1503pp; English.

XX

CC

The invention relates to 566 new isolated human polypeptides and their

CC

encoding genes, sequences that are at least 95% identical to these or

CC

sequences comprising one or more conservative substitutions in these. The

CC

polypeptide, polynucleotide and antibodies against the polypeptides are

CC

useful in diagnosing, treating or preventing NOVX-associated disorders,

CC

e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,

CC

diabetes, AIDS, multiple sclerosis, graft-versus-host disease,

CC

Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.

CC

The nucleic acids are further used as hybridization probes, in chromosome

CC

mapping, tissue typing, preventive medicine, and pharmacogenomics. The

CC

polypeptides are also useful as vaccines. This sequence represents an

CC

example of the polypeptide of the invention.

XX

SQ

Sequence 851 AA;

Query Match 37.4%; Score 4482; DB 8; Length 851;

Best Local Similarity 99.8%; Pred. No. 1.1e-261; Mismatches 2; Indels 0; Gaps 0;

Matches 849; Conservative 0;

QY 1053 PEGFVGNLTYESISSTAINVSWPPAQNGLVYVYVSLILQQTPRHVRPLVYERSIYF 1112

DB 1 PEGFVGNLTYESISSTAINVSWPPAQNGLVYVYVSLILQQTPRHVRPLVYERSIYF 60

QY 1113 DNLEKYDYILKTPSTKGFSDTYTAQLYIKTEEDVPTSPINTFKNLSSTVLLSWD 1172

DB 61 DNLEKYDYILKTPSTKGFSDTYTAQLYIKTEEDVPTSPINTFKNLSSTVLLSWD 120

QY 1173 PPVKNGAIIISYDLTLQGNENYSFITSNDYIIIEELSPTLYSFFAAARTRKGLGPSSI 1232

DB 121 PPVKNGAIIISYDLTLQGNENYSFITSNDYIIIEELSPTLYSFFAAARTRKGLGPSSI 180

QY 1233 LFFYTDSEVPLAPPQNLTINCTDFVWLKNPSPPLGPGIVKVYSFKIHEHEDTIIYKN 1292

DB 181 LFFYTDSEVPLAPPQNLTINCTDFVWLKNPSPPLGPGIVKVYSFKIHEHEDTIIYKN 240

QY 1293 ISGFKTEAKLVGLEPVSITYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMOCWATS 1352

DB 241 ISGFKTEAKLVGLEPVSITYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMOCWATS 300

QY 1353 WQSVLVKWDPPPKANGIITQYMTVERNSTKVSPODHMTFFIKLLANTSYVFKVRASTSA 1412

DB 301 WQSVLVKWDPPPKANGIITQYMTVERNSTKVSPODHMTFFIKLLANTSYVFKVRASTSA 360

QY 1413 GEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTCOLRAQ 1472

DB 361 GEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTCOLRAQ 420

QY 1473 KCKEWESECEVYQKIQLIYEAHLTEETVYGLKFRWYRFOVAASTNAGYNASWISTK 1532

Db	421	KCKEWESECVYQIKYLYEAHLTEETVYGLKFRWYRFQVAASNAGYNASNIWK	480
Qy	1533	TLPGPDGPPENHVATSPFSISISWSSEPAVITGPTCYLIDVKSVDNDEFNISFIKSN	1592
Db	481	TLPGPDGPPENHVATSPFSISISWSSEPAVITGPTCYLIDVKSVDNDEFNISFIKSN	540
Qy	1593	ENKTIKDLFTFRTYVSVITFTGNISAAVYVEGSSAEMIVTTLESAPKDPNNMTFOK	1652
Db	541	ENKTIKDLFTFRTYVSVITFTGNISAAVYVEGSSAEMIVTTLESAPKDPNNMTFOK	600
Qy	1653	IPDEVTKFOLTPSPQNGNIQVQALVYRDDPTAVQIHNLISIIQKTNFTVIAMLEGL	1712
Db	601	IPDEVTKFOLTPSPQNGNIQVQALVYRDDPTAVQIHNLISIIQKTNFTVIAMLEGL	660
Qy	1713	KGHTYNTISVYAVNSAGAPKVPMTITMDIKAPAPKTKPTPIYDATGKLLVTSITIR	1772
Db	661	KGHTYNTISVYAVNSAGAPKVPMTITMDIKAPAPKTKPTPIYDATGKLLVTSITIR	720
Qy	1773	MEICYSDDHGPINKVQVLATETGAQHDGNVTWKVDAYFNKARPFTWEGFPNPPCTEGK	1832
Db	721	MEICYSDDHGPINKVQVLATETGAQHDGNVTWKVDAYFNKARPFTWEGFPNPPCTEGK	780
Qy	1833	TKPSGNEEYIIGADNACMIPGNEKICNGPLKPKKQYLFKPRATNIMGQFTDSYSDPV	1892
Db	781	TKPSGNEEYIIGADNACMIPGNEKICNGPLKPKKQYLFKPRATNIMGQFTDSYSDPV	840
Qy	1893	KTLGEGLSERT	1903
Db	841	KTLGEGLSERT	851
RESULT 15			
ADH41617	ID ADH41617 standard; protein; 740 AA.		
XX	AC	ADH41617;	
XX	DT	25-MAR-2004 (first entry)	
XX	DE	Novel human protein NOV15c.	
KW	cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;		
KW	antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;		
KW	antiparkinsonian; antiaethmatic; antiinfertility; cardiomyopathy;		
KW	atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;		
KW	multiple sclerosis; graft-versus-host disease; Alzheimer's disease;		
KW	Parkinson's disease; asthma; fertility disorder; chromosome mapping;		
KW	tissue typing; preventive medicine; pharmacogenomic; vaccine.		
OS	Homo sapiens.		
XX	PN	WO2003.102159-A2.	
XX	PD	11-DEC-2003.	
XX	PF	04-JUN-2003; 2003WO-US017573.	
XX	PR	04-JUN-2002; 2002US-0385490P.	
PR	04-JUN-2002; 2002US-0385615P.		
PR	04-JUN-2002; 2002US-0385755P.		
PR	05-JUN-2002; 2002US-0386041P.		
PR	06-JUN-2002; 2002US-0386355P.		
PR	06-JUN-2002; 2002US-0386357P.		
PR	06-JUN-2002; 2002US-0386447P.		
PR	06-JUN-2002; 2002US-0386459P.		
PR	06-JUN-2002; 2002US-0386465P.		
PR	07-JUN-2002; 2002US-0386701P.		
PR	07-JUN-2002; 2002US-0386796P.		
PR	07-JUN-2002; 2002US-0386931P.		
PR	07-JUN-2002; 2002US-0387078P.		
PR	07-JUN-2002; 2002US-0387081P.		

PR	07-JUN-2002; 2002US-0387083P.		
PR	10-JUN-2002; 2002US-0387429P.		
PR	10-JUN-2002; 2002US-0387540P.		
PR	10-JUN-2002; 2002US-0387666P.		
PR	11-JUN-2002; 2002US-0387666P.		
PR	11-JUN-2002; 2002US-0387610P.		
PR	11-JUN-2002; 2002US-0387659P.		
PR	11-JUN-2002; 2002US-0387668P.		
PR	11-JUN-2002; 2002US-0387696P.		
PR	11-JUN-2002; 2002US-0387859P.		
PR	12-JUN-2002; 2002US-0387934P.		
PR	12-JUN-2002; 2002US-0387960P.		
PR	12-JUN-2002; 2002US-0388022P.		
PR	12-JUN-2002; 2002US-0388096P.		
PR	12-JUN-2002; 2002US-0388432P.		
PR	12-JUN-2002; 2002US-0388479P.		
PR	13-JUN-2002; 2002US-0389123P.		
PR	14-JUN-2002; 2002US-0389120P.		
PR	14-JUN-2002; 2002US-0389146P.		
PR	17-JUN-2002; 2002US-0389742P.		
PR	18-JUN-2002; 2002US-0389604P.		
PR	18-JUN-2002; 2002US-0389884P.		
PR	19-JUN-2002; 2002US-0390008P.		
PR	19-JUN-2002; 2002US-0390144P.		
PR	19-JUN-2002; 2002US-0390209P.		
PR	25-JUN-2002; 2002US-0391726P.		
PR	06-AUG-2002; 2002US-0401628P.		
PR	09-AUG-2002; 2002US-0402268P.		
PR	12-AUG-2002; 2002US-0402822P.		
PR	13-AUG-2002; 2002US-0403458P.		
PR	15-AUG-2002; 2002US-0403617P.		
PR	15-AUG-2002; 2002US-0403732P.		
PR	26-AUG-2002; 2002US-0406182P.		
PR	12-SEP-2002; 2002US-0410085P.		
PR	13-SEP-2002; 2002US-0410505P.		
PR	23-SEP-2002; 2002US-0412955P.		
PR	30-SEP-2002; 2002US-0415195P.		
PR	23-OCT-2002; 2002US-0420627P.		
PR	23-OCT-2002; 2002US-0420718P.		
PR	24-OCT-2002; 2002US-0420852P.		
PR	31-OCT-2002; 2002US-0422750P.		
PR	01-NOV-2002; 2002US-0423095P.		
PR	05-NOV-2002; 2002US-0423748P.		
XX	(CURA-) CURAGEN CORP.		
PI	Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;		
PI	Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;		
PI	Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;		
PI	Grosse WM, Gunther E, Guo X, Gusev VY, Hertrmann JL, Ji W, Kekuda R;		
PI	Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;		
PI	MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;		
PI	Milliet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;		
PI	Rieger DK, Rothenberg ME, Shenoy SG, Shinkets RA, Smithson G;		
PI	Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;		
PI	Wolenc AR, Zhong M, Zhong H;		
XX	WPI; 2004-053467/05.		
DR	N-PSDB; ADH41616.		
XX	New NOVX polypeptides and nucleic acid molecules useful for preventing or		
PT	treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,		
PT	atherosclerosis or diabetes, in chromosome mapping, tissue typing or in		
PT	pharmacogenomics.		
XX	Claim 2; SEQ ID NO 170; 1503pp; English.		
PS	The invention relates to 566 new isolated human polypeptides and their		
CC	encoding genes, sequences that are at least 95% identical to these or		
CC	sequences comprising one or more conservative substitutions in these. The		
CC	polypeptide, polynucleotide and antibodies against the polypeptides are		
CC	useful in diagnosing, treating or preventing NOVX-associated disorders,		
CC	e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity.		

CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the polypeptide of the invention.
XX
SQ

Sequence 740 AA;

Query Match 31.7%; Score 3804; DB 8; Length 740;

Best Local Similarity 99.7%; Pred. No. 8.4e-221;

Matches 734; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	305	PEGPPQNCVTGNTGKSFSLWDPPPIVTKGFSYRVELYGPGRILDNSTKDLKFAFTNL	364
Db	3	PEGPPQNCVTGNTGKSFSLWDPPPIVTKGFSYRVELYGPGRILDNSTKDLKFAFTNL	62
QY	365	TPFTMDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQAEVSTQVRIITWKKPRQPN	424
Db	63	TPFTMDVYIAAETSAGTGPKSNISVFTPPDVGAVFDLQAEVSTQVRIITWKKPRQPN	122
QY	425	GIINQVRVKVLVPETGIIILENTLLTGNEYINDPMAPEIVNIPEPMVGLYEGSAEMSSDL	484
Db	123	GIINQVRVKVLVPETGIIILENTLLTGNEYINDPMAPEIVNIPEPMVGLYEGSAEMSSDL	182
QY	485	HSLATFIYNHSHDKPNPARNRABDQTSVVVTRNQVITDIAAQQLSVIRRLVPFTEHMI	544
Db	183	HSLATFIYNHSHDKPNPARNRABDQTSVVVTRNQVITDIAAQQLSVIRRLVPFTEHMI	242
QY	545	SVSAPTIMGEGPTVLVSRTRQOVPSIKIINYNKISSSSILLYWDPPEYPNGKITHYTI	604
Db	243	SVSAPTIMGEGPTVLVSRTRQOVPSIKIINYNKISSSSILLYWDPPEYPNGKITHYTI	302
QY	605	YAMELDTNRAFOITTDNSFLITGLKKYTKYKRVAASTHGGSSLSSENDIFVRTSEDE	664
Db	303	YAMELDTNRAFOITTDNSFLITGLKKYTKYKRVAASTHGGSSLSSENDIFVRTSEDE	362
QY	665	PESSPDQVEIDVTADRIKWSPEKPNGLIAYEVLYKNIDTLYMKNSTTTDIIIRNL	724
Db	363	PESSPDQVEIDVTADRIKWSPEKPNGLIAYEVLYKNIDTLYMKNSTTTDIIIRNL	422
QY	725	RPHLYNISVRSYTRFGHGNQVSSLSVRSSETVPSAPENITYKNISSGEIELSLFPPS	784
Db	423	RPHLYNISVRSYTRFGHGNQVSSLSVRSSETVPSAPENITYKNISSGEIELSLFPPS	482
QY	785	SPNGIIKKYTYILKRSNGNEERTINTSLTQNIKVLKXYQYIIIEVSASTLKGEGVRSAP	844
Db	483	SPNGIIKKYTYILKRSNGNEERTINTSLTQNIKVLKXYQYIIIEVSASTLKGEGVRSAP	542
QY	845	ISILTEADAPDSPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSSLKTINVT	904
Db	543	ISILTEADAPDSPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSSLKTINVT	602
QY	905	ETSLEISDLDDYNVEYSAYVTASTRFGDKTGSNIISFQTEGAPSDPPKDVYVYANLSSSS	964
Db	603	ETSLEISDLDDYNVEYSAYVTASTRFGDKTGSNIISFQTEGAPSDPPKDVYVYANLSSSS	662
QY	965	IILFWTPPSKPGNGIIOYYSVYVNTSGTFMONTLHETLNDNDNMVSTIIDKLTIPSY	1024
Db	663	IILFWTPPSKPGNGIIOYYSVYVNTSGTFMONTLHETLNDNDNMVSTIIDKLTIPSY	722
QY	1025	TFWLTASTSVGNGNKS 1040	
Db	723	TFWLTASTSVGNGNKS 738	

Search completed: June 30, 2005, 14:59:18

Job time : 249 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:54:09 ; Search time 58 Seconds
(without alignments)
2948.638 Million cell updates/sec

Title: US-10-673-885-2
Perfect score: 11985
Sequence: 1 MDLFIPLLLFGTSETQVD.....MDSLDMEGDVLEWEETM 2291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11985	100.0	2291	4	US-09-822-871-2
2	10536	87.9	2301	4	US-09-822-871-4
3	4535.5	37.8	898	1	US-08-036-210-22
4	4535.5	37.8	898	2	US-08-449-609-22
5	4535.5	37.8	898	4	US-09-361-096A-22
6	1903.5	15.9	402	1	US-08-036-210-15
7	1903.5	15.9	402	2	US-08-449-609-15
8	1902.5	15.9	401	4	US-09-361-096A-15
9	1713.5	14.3	322	1	US-08-036-210-11
10	1713.5	14.3	322	2	US-08-449-609-11
11	1713.5	14.3	322	4	US-09-361-096A-11
12	1538.5	12.8	289	4	US-09-361-096A-47
13	1194.5	10.0	334	1	US-08-036-210-7
14	1194.5	10.0	334	2	US-08-449-609-7
15	1194.5	10.0	334	4	US-09-361-096A-7
16	1154	9.6	1911	1	US-08-800-825A-5
17	1154	9.6	1911	2	US-08-800-825A-5
18	1154	9.6	1911	3	US-09-158-657-5
19	1154	9.6	1911	5	PCT-US94-10166-5
20	1107	9.2	1997	4	US-09-949-016-6275
21	1084.5	9.0	1903	4	US-09-949-016-8049
22	942	7.9	176	1	US-08-036-210-9
23	942	7.9	176	2	US-08-449-609-9
24	942	7.9	176	4	US-09-361-096A-9
25	836	7.0	1188	1	US-08-201-697-4
26	834.5	7.0	1246	4	US-09-949-016-8051
27	834.5	7.0	1246	4	US-09-949-016-8052

28	834.5	7.0	1246	4	US-09-949-016-8053	Sequence 8053, Ap
29	834.5	7.0	1246	4	US-09-949-016-8054	Sequence 8054, Ap
30	834.5	7.0	1246	4	US-09-949-016-8055	Sequence 8055, Ap
31	833	7.0	1501	2	US-08-447-464-3	Sequence 3, Appli
32	833	7.0	1501	2	US-08-716-679-3	Sequence 3, Appli
33	827.5	6.9	1187	1	US-08-201-697-2	Sequence 2, Appli
34	823.5	6.9	1274	4	US-09-949-016-8828	Sequence 8828, Ap
35	823.5	6.9	1274	4	US-09-949-016-8829	Sequence 8829, Ap
36	823.5	6.9	1274	4	US-09-949-016-8830	Sequence 8830, Ap
37	823.5	6.9	1274	4	US-09-949-016-8831	Sequence 8831, Ap
38	823.5	6.9	1274	4	US-09-949-016-8832	Sequence 8832, Ap
39	820	6.8	1337	3	US-08-854-585-2	Sequence 2, Appli
40	820	6.8	1337	4	US-09-447-533-2	Sequence 2, Appli
41	820	6.8	1337	5	PCT-US95-05512-2	Sequence 2, Appli
42	746.5	6.2	1711	2	US-08-342-930-2	Sequence 2, Appli
43	740.5	6.2	1452	2	US-08-449-644-8	Sequence 8, Appli
44	740.5	6.2	1452	2	US-08-087-244A-8	Sequence 8, Appli
45	740	6.2	506	4	US-09-949-016-8833	Sequence 8833, Ap

ALIGNMENTS

RESULT 1
US-09-822-871-2
; Sequence 2, Application US/09822871
; Patent No. 6723547
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219
; CURRENT APPLICATION NUMBER: US/09/822,871
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2291
; TYPE: PRT
; ORGANISM: Human
US-09-822-871-2

Query Match		100.0%;	Score 11985;	DB 4;	Length 2291;
Best Local Similarity		100.0%;	Pred. NO. 0;		
Matches 2291;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MDFLIPLLLFGTSETQVDVSNVVPGRYDITISISTYTSPTVTRIVTPNVTKGPPV	60		
DB	1	MDFLIPLLLFGTSETQVDVSNVVPGRYDITISISTYTSPTVTRIVTPNVTKGPPV	60		
QY	61	FLAGERVGSAGILLSNWTPPNNGRIISIVKYKEVCPMMQTVYTVQVRSKPSLEVLTTN	120		
DB	61	FLAGERVGSAGILLSNWTPPNNGRIISIVKYKEVCPMMQTVYTVQVRSKPSLEVLTTN	120		
QY	121	LNPGTTYEIKVAENSAGIGVSDPFLQATASAFKGVNLTVEAYNASAVKLIWYLPKQ	180		
DB	121	LNPGTTYEIKVAENSAGIGVSDPFLQATASAFKGVNLTVEAYNASAVKLIWYLPKQ	180		
QY	181	PNGKITSPKISVKHARSGIVKDVIRVEDILTGLKPECNENSESFLWSTASPSPTLGRV	240		
DB	181	PNGKITSPKISVKHARSGIVKDVIRVEDILTGLKPECNENSESFLWSTASPSPTLGRV	240		
QY	241	TPPSRTTHSSSTLTONEISSVWKEPISFVWTHLRPVYTVLFEVSAATTEAGYIDSTIVRT	300		
DB	241	TPPSRTTHSSSTLTONEISSVWKEPISFVWTHLRPVYTVLFEVSAATTEAGYIDSTIVRT	300		
QY	301	PESVPEGPQNCVGNITGKSFISILDWDPPTIVTGKFSVRVELYGPSGRILDNSTKDKFA	360		
DB	301	PESVPEGPQNCVGNITGKSFISILDWDPPTIVTGKFSVRVELYGPSGRILDNSTKDKFA	360		
QY	361	FTNLTPTMYDVYIAAETSAGTGPKNISVFTTTPVPGAVFDQLAEVESTQVITWKKP	420		

Db 361 FTNLPTFTMYDVYIAAETTSAGTCKPSKSNISVFTPPDVPAGVFDQLQAEVESTQVRIWTKKP 420
Qy 421 ROPNGLIINQYRKVLVPETGIIILENTLLTGNEYNINDPMAPELVNIIVBPMVGLYEGSAEM 480
Db 421 ROPNGLIINQYRKVLVPETGIIILENTLLTGNEYNINDPMAPELVNIIVBPMVGLYEGSAEM 480
Qy 481 SSDLSLAFIYNINSHDPKFPARNRAEDQTSVVTTRNQYITDIAAQSLSYVIRRLVPPT 540
Db 481 SSDLSLAFIYNINSHDPKFPARNRAEDQTSVVTTRNQYITDIAAQSLSYVIRRLVPPT 540
Qy 541 EHMISVSAFTINGEGPPTVLSVRTRQVPSSIKIINYKNISSSILLYWDPPPEYNGKIT 600
Db 541 EHMISVSAFTINGEGPPTVLSVRTRQVPSSIKIINYKNISSSILLYWDPPPEYNGKIT 600
Qy 601 HTYIYAMELDTNRAFOITIDNSFLITGLKKYTKYKORVAASTHOGESSLSSENDIFVRT 660
Db 601 HTYIYAMELDTNRAFOITIDNSFLITGLKKYTKYKORVAASTHOGESSLSSENDIFVRT 660
Qy 661 SEDEPRESSQDVEVDVTADEIRLKWSPPEKNGIIIAEVLVYKNIDTLYMKNTSTTDII 720
Db 661 SEDEPRESSQDVEVDVTADEIRLKWSPPEKNGIIIAEVLVYKNIDTLYMKNTSTTDII 720
Qy 721 LRNLRPHTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSF 780
Db 721 LRNLRPHTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSF 780
Qy 781 LPPSSNGIIKKYTIILKESNGNEERTINTSLTQNIKVKKYQYIIIVSASTLKGEVY 840
Db 781 LPPSSNGIIKKYTIILKESNGNEERTINTSLTQNIKVKKYQYIIIVSASTLKGEVY 840
Qy 841 RGAPTSILTEADAPSPDQFSVKQLSGVTVKLSWQPPLEPNGLIILYYTVVWNRSSSLKT 900
Db 841 RGAPTSILTEADAPSPDQFSVKQLSGVTVKLSWQPPLEPNGLIILYYTVVWNRSSSLKT 900
Qy 901 INVTTLSLSLDLYNVVEYSAYVASTRFGDGKTSNIIISFQTPPEGAPSDPPKQVYVYANL 960
Db 901 INVTTLSLSLDLYNVVEYSAYVASTRFGDGKTSNIIISFQTPPEGAPSDPPKQVYVYANL 960
Qy 961 SSSSIIILFWTPSPKNGIIQYYSVYRNSTSGTFMQNFTLHETLNDPDMNTVSTIIDKLT 1020
Db 961 SSSSIIILFWTPSPKNGIIQYYSVYRNSTSGTFMQNFTLHETLNDPDMNTVSTIIDKLT 1020
Qy 1021 FSYTFEWTASTSVGNKSSDIIIEVYTQDIPGEGVGNLTYESISSTAINVSWPPAPQ 1080
Db 1021 FSYTFEWTASTSVGNKSSDIIIEVYTQDIPGEGVGNLTYESISSTAINVSWPPAPQ 1080
Qy 1081 NGLVFFVYVSLIILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSDTYTAQ 1140
Db 1081 NGLVFFVYVSLIILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTKGFSDTYTAQ 1140
Qy 1141 LYIKTEEDVPETSPINTFNKLSSTSVLLSWDPPVKPNCALISYDILTLOGPNEYSFITS 1200
Db 1141 LYIKTEEDVPETSPINTFNKLSSTSVLLSWDPPVKPNCALISYDILTLOGPNEYSFITS 1200
Qy 1201 DNVIIIEELSPPFLYSFFFAAARTRKGLGPSSILFFVYTDSESVPLAPPQNLTLINCTSDFW 1260
Db 1201 DNVIIIEELSPPFLYSFFFAAARTRKGLGPSSILFFVYTDSESVPLAPPQNLTLINCTSDFW 1260
Qy 1261 LKMWSPSPFGGIYKVKYSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSSTYIRSAPTK 1320
Db 1261 LKMWSPSPFGGIYKVKYSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSSTYIRSAPTK 1320
Qy 1321 VGNQGNFNVVKTTOESVDPVQVQNMCHATSQSVLVKWDPPKANGIIITQYMTVERN 1380
Db 1321 VGNQGNFNVVKTTOESVDPVQVQNMCHATSQSVLVKWDPPKANGIIITQYMTVERN 1380
Qy 1381 STKVSQDQDMYTFIKLANTSYVFKVRASTAGDESTCHVSTLTPETVPSPTNIAFSD 1440
Db 1381 STKVSQDQDMYTFIKLANTSYVFKVRASTAGDESTCHVSTLTPETVPSPTNIAFSD 1440
Qy 1441 VQSTSATLWIRPDTILGYFQNYKIITQLRAQCKEWESECEVYQKIQYLYEAHLTEET 1500
Db 1441 VQSTSATLWIRPDTILGYFQNYKIITQLRAQCKEWESECEVYQKIQYLYEAHLTEET 1500

Qy 1501 VYGLKKFRWYRFOVAASTNAGYGNASNWISTKTLPGPDGPPENVHVVATSPFSISISWS 1560
Db 1501 VYGLKKFRWYRFOVAASTNAGYGNASNWISTKTLPGPDGPPENVHVVATSPFSISISWS 1560
Qy 1561 EPAVITGPTCYLIDVKSVDNDEFNISFKISNEENKTIKIDLEIPTRYSVWITAFGTNIS 1620
Db 1561 EPAVITGPTCYLIDVKSVDNDEFNISFKISNEENKTIKIDLEIPTRYSVWITAFGTNIS 1620
Qy 1621 AAYVEGKSAEMIVTTLSAPKDPNNMTFOKIPDEVTKFQTLFPLPPSPNGNIQVYQAL 1680
Db 1621 AAYVEGKSAEMIVTTLSAPKDPNNMTFOKIPDEVTKFQTLFPLPPSPNGNIQVYQAL 1680
Qy 1681 VYREDPTAVQIHNLSIIQKNTFVIAMLEGLKGGHTYINISVYVNSAGAGKVPWRITM 1740
Db 1681 VYREDPTAVQIHNLSIIQKNTFVIAMLEGLKGGHTYINISVYVNSAGAGKVPWRITM 1740
Qy 1741 DIKAPARPKTPTIYDATGKLLVTSTTITIRMPICYSDDDHGPICKNVQVLTATETGAQHD 1800
Db 1741 DIKAPARPKTPTIYDATGKLLVTSTTITIRMPICYSDDDHGPICKNVQVLTATETGAQHD 1800
Qy 1801 GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIIGADNACMIPGHNEDKIC 1860
Db 1801 GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNEEYIIIGADNACMIPGHNEDKIC 1860
Qy 1861 NGPLPKKQYLPKFRATNIMGQFTDSDYSDPVKTLGEGLSERTVEIILSVTLCLISILL 1920
Db 1861 NGPLPKKQYLPKFRATNIMGQFTDSDYSDPVKTLGEGLSERTVEIILSVTLCLISILL 1920
Qy 1921 GTAIFAFARIROKQEGGTYSPQDAEIIDTKLKLQDLITVADLELKDRLTRPISKKSPL 1980
Db 1921 GTAIFAFARIROKQEGGTYSPQDAEIIDTKLKLQDLITVADLELKDRLTRPISKKSPL 1980
Qy 1981 QHVEBELCTNNLKFQBESELSKFLQDLSSSTDADLPWNRKORFPNKPNNNNRVKLI 2040
Db 1981 QHVEBELCTNNLKFQBESELSKFLQDLSSSTDADLPWNRKORFPNKPNNNNRVKLI 2040
Qy 2041 DASVPGSDVINASYISGVLCNPEFIATQGLPCTVGDVFWVWVWETRAKTLVMLTQCFEKG 2100
Db 2041 DASVPGSDVINASYISGVLCNPEFIATQGLPCTVGDVFWVWVWETRAKTLVMLTQCFEKG 2100
Qy 2101 RIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRCNFTAWPE 2160
Db 2101 RIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRCNFTAWPE 2160
Qy 2161 HGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTHOHINDHDFVDI 2220
Db 2161 HGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTHOHINDHDFVDI 2220
Qy 2221 YGLVABLRSERCMQVONLAQYIFLHQCLDLSNKGSNQPICFVNVYSALQKMDSLDAMEG 2280
Db 2221 YGLVABLRSERCMQVONLAQYIFLHQCLDLSNKGSNQPICFVNVYSALQKMDSLDAMEG 2280
Qy 2281 DVELEWEETM 2291
Db 2281 DVELEWEETM 2291

RESULT 2

US-09-822-871-4
; Sequence 4, Application US/09822871
; Patent No. 6723547
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219
; CURRENT APPLICATION NUMBER: US/09/822, 871
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4

```

; LENGTH: 2301
; TYPE: PRT
; ORGANISM: Rat
US-09-822-871-4

Query Match      87.9%; Score 10536; DB 4; Length 2301;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

QY 1 MDPLIIFLLFICTSETQVDNSNVPGTRIDITISS-ITTTSPVTRIVTPNVTXPGPP 59
DB 1 MDPHRSFLLLIGTSSQVDVSSPDGTGVDITLSSVSATTVSSPVSRITLATNVTXPGPP 60
QY 60 VFLAGRVCSAGILLSWNTPPNPGRHIIISYIVKYKEVCPWMQTVYTVQVRSKDSLEVLIT 119
DB 61 VFLAGRVCSAGILLSWNTPPNPGRHIIISYIVKYKEVCPWMQTVYTVQVRSKDSLEVLIT 120
QY 120 NLNPGTYYEIKVAEENSAGIGVPSDFPQTAESAPGVNLTVEAYNASAVKLTIWYLR 179
DB 121 NLNPGTYYEIKVAEENSAGIGVPSDFPQTAESAPGVNLTVEALNYSANVLTIWYLR 180
QY 180 QNGKITSPKISVKHARSIGVVKDVSIRVEDIITGKLPECNENSSESFLWSTASPSPTLGR 239
DB 181 QNGKITSPKISVKHARSIGVVKDVSIRVEDIITGKLPECNENSSESFLWSTASPSPTLGR 240
QY 240 VTPPSRTTHSSSTLTONEISSVWKEPISFVWTHLRPYTYLFEVSAATTEAGYIDSTIVR 299
DB 241 VTPVTRTQSSSTAARSKISSVWKEPISFVWTHLRPYTYLFEVSAATTEAGYIDSTIVR 300
QY 300 TPESVPEGPONCVTCNITCKFSILWDPTTIVTGKFSYRVELYGPSSGRIILNYSKOLXP 359
DB 301 TPESVPEGPONCIMGNTGKAFSISWDPTTIVTGKFSYRVELYGPSSGRIILNYSKOLXP 360
QY 360 AFTNLTPFTMYDVYIAAETSAGTGPKNISVFTPPDVGAVFDLQLAEVESQVRLTWCK 419
DB 361 AFTNLTPFTMYDVYIAAETSAGTGPKNISVFTPPDVGAVFDLQLAEVESQVRLTWCK 420
QY 420 PRQNGIINQYRVKVLVPGTGIIENLTLLTGNNEYINDPMAPEINIVBPMVGLYEGSAB 479
DB 421 PRQNGIISQYRVKVSLETGVVLENTLLTGQDESINPMSPREIMNLVDPMGIFEGSGE 480
QY 480 MSSDLHSLATFYNSHPDKNPARNAEDQTPVVTTRNQYITDIAEOLSVIRRLVPP 539
DB 481 MSSDLHSPAFYNSHPNDFFASTRAEQSSPVVTTRNQYITDIAEOLSVIRRLVPP 540
QY 540 TEHMSVSAFTIMGEGPPTVLSVTRTEQVPSIIKINYNKNISSSILLWDPPEYNGKI 599
DB 541 TEHMSVSAFTIMGEGPPTVLSVTRTEQVPSIIKINYNKNISSSILLWDPPEYNGKI 600
QY 600 THYTIYAMBLDINRAFOITTDNSFLITGLKKYTKYKMRVAASHTDGSSESLSEENDIFVR 659
DB 601 THYTIYATELDINRAFOITTDNSFLITGLKKYTKYKMRVAASHTDGSSESLSEENDIFVR 660
QY 660 TSEDEPSSQDVEVDVTADEIRLKWSPPEKNGIIIAYEVLVYKXNIDFLYKNTSTTDI 719
DB 661 TSEDEPSSQDQVQTVGSFSELRLKWSPPEKNGIIIAYEVLVYKXNADTLFVKNSTTDI 720
QY 720 ILNRLAPHTLYNISVRSYTRFGHGNQVSSILSVRTSETVPDSAPENITTKNISSGEIELS 779
DB 721 IISDLKPXYLNISSYTRLGHGNQVSSILSVRTSETVPDSAPENITTKNISSGEIELS 780
QY 780 FLPPSPNGIISKYTYLKRSGNEERTINTTSLQNIKVLYKQYIIEVSASTLKGRG 839
DB 781 FLPPSPNGIISKYTYLKRSGNEERTINTTSLQNIKVLYKQYIIEVSASTLKGRG 840
QY 840 VRSAPISILTEADAPSPQDPSVKOLSGVTVKLSWQPLEPENGIIILYTVVMNRSLX 899
DB 841 IRSRPSILTEADAPSPQDPSVKOLSGVTVKLSWQPLEPENGIIILYTVVMNRSLX 900
QY 900 TINVTSTLESLDLDYNNYSAVYASTRFGDGKTCNSIISQTPBGPSPDPKQVYAN 959
DB 901 AINATEASLVSLDLDYNNYSAVYASTRFGDGKTCNSIISQTPBGPSPDPKQVYAN 960

960 LSSSSIIILFWTPPSKNGIIQYYSVYRYNTSCTFMQNTLHELHETLNDPQNTVSTIIDKLT 1019
961 LSSSSIIILFWTPPSKNGIIQYYSVYRYNTSCTFMQNTLHELHETLNDPQNTVSTIIDKLT 1020
1020 IFSYTYFWLTASTSVGNKSSDIIIEVYTDQDIPGFGVGNLTYSSTAIINVSVPVPAQ 1079
1021 IFSYTYFWLTASTSVGNKSSDIIIEVYTDQDIPGFGVGNLTYSSTAIINVSVPVPAQ 1080
1080 PNGLYFYVSLIQT-PRHVRPPLVTVYRSYIFONLEKYTDIILKITPSTKSGSDTYT 1138
1081 PNGLYFYVSLIQT-PRHVRPPLVTVYRSYIFONLEKYTDIILKITPSTKSGSDTYT 1140
1139 AQLYKTEEDVETSPITINTFNKLSSTVLSLWSDPVPKNGAIIISDLTLQGNENYSFI 1198
1141 TQHIKTEEDVETSPITINTFNKLSSTVLSLWSDPVPKNGAIIISDLTLQGNENYSFI 1200
1199 TSDNIIILBELSPFLYSFFAAARTRKGLGPSILFFYTDSEVPLAPPONLTLINCTSD 1258
1201 TSGNHIVLEELSPFLYSFFAAARTRKGLGPSILFFYTDSEVPLAPPONLTLINCTSD 1260
1259 VMLKWSPLPGGIYKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRVSAP 1318
1261 VMLTWSPLPGGIYKVSFKIHEHETDTIYKNIISGFKTEAKLVGLEPVSYSIRVSAP 1320
1319 TKVGNQPSNVKTTQESVDPVQVQNMCMATSHQSVLVKWDPPKPKANGIITQYMYTVS 1378
1321 TKVGNQPSNVKTTQESVDPVQVQNMCMATSHQSVLVKWDPPKPKANGIITQYMYTVS 1380
1379 RNSTKVSQDQMYTITKLLANTSYVFKVRASTAGEDESTCHVSTLPTVPSVPTNIAF 1438
1381 GNSTKVSQDQMYTITKLLANTSYVFKVRASTAGEDESTCHVSTLPTVPSVPTNIAF 1440
1439 SDVQSTSATLWIRPDTILGPFQNYKITTLQRAQCKEWESECEVEYQKIQYLYEAHLE 1498
1441 SNVQSTSATLWIRPDTILGPFQNYKITTLQRAQCKEWESECEVEYQKIQYLYEAHLE 1500
1499 ETVYGLKPRWYRFOVAASTNAGYGNASNIISTKTLPGPPDPPENHVHVSATSPSIS 1558
1501 ETVYGLKPRWYRFOVAASTNAGYGNASNIISTKTLPGPPDPPENHVHVSATSPSIS 1560
1559 RSEPAVITGPTCYLLDKVSDNDENFIISFLKSNEENKTEIKDLAIFRYSVIVTAFTCN 1618
1561 RSEPAVITGPTCYLLDKVSDNDENFIISFLKSNEENKTEIKDLAIFRYSVIVTAFTCN 1620
1619 ISAAVEGKSAAEMVTTLLESAPDPPNNMTFQKIPDEVTKEFQLTFLPPSQPNQNIQVQ 1678
1621 VSRATYDCKSSAEVITTLLESVQKPPNNMTFQKIPDEVTKEFQLTFLPPSQPNQNIQVQ 1680
1679 ALVYREDDPATAVOIHNLISIIQKNTFTVIAMLEGLKGGHTYNIISVAVNSAGAPKVPMRI 1738
1681 ALVYREDDPATAVOIHNLISIIQKNTFTVIAMLEGLKGGHTYNIISVAVNSAGAPKVPMRI 1740
1739 TMDIKAPARKPTPIVDATGKLVSTTITIRMPICYSDHGHPIKXVOVLATETGAQ 1798
1741 TMDIKAPARKPTPIVDATGKLVSTTITIRMPICYSDHGHPIKXVOVLATETGAQ 1800
1799 HDGNTVKWYDAVFNKARPYFTNEGPNPCEGKTKFSGNEEYIIGADNACWIPGNEOK 1858
1801 QDGNVTKWYDAVFNKARPYFTNEGPNPCEGKTKFSGNEEYIIGADNACWIPGNEOK 1860
1859 ICNGPLKPKQYLFKFRATNIMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCILSII 1918
1861 ICNGPLKPKQYLFKFRATNIMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCILSII 1920
1919 LLGTAIPAPARIQKQEGGTYSPDASIIIDTKLQDLITVADLELKDRLTR----- 1972
1921 LLGTAIPAPARIQKQEGGTYSPDASIIIDTKLQDLITVADLELKDRLTR----- 1980
1973 ---PISKSKFLQHVLEELCTNNNLKQEFSELPKFLQDLSSTDAOLPNNRKNRPNIKP 2029
1981 STKPSKSKFLQHVLEELCTNNNLKQEFSELPKFLQDLSSTDAOLPNNRKNRPNIKP 2040
2030 YNNNRVKLADASVFGSDYINASYISGLPCNNEFIATQGPLGTGDFWRMVWETRAKT 2089
```

Db 2041 Y-NNNRVRLIADVSLPGSDYINASYVSGYLCNEFIATQGPLPGTVGDFWRMWRTRTKT 2099
QY 2090 LVMLTQCFKGRIRCHOYWPEDNKPVTYVGDIVITKLMEDVDQIDWTIRDLKIERHGDQMT 2149
Db 2100 LVMLTQCFKGRIRCHOYWPEDNKPVTYVGDIVITKLMEDIQIDWTIRDLKIERHGDQMT 2159
QY 2150 VRQCNFTAPHEGVSPNSAPLHFVKLVRASRAHDTTPMIVHCSAGVGRGVFIADHLT 2209
Db 2160 VRQCNFTGPEKGVPENTTFLHFVKLVTSRAHDTTPMVHCSAGVGRGVFIADHLT 2219
QY 2210 QHINHDVFDIYGLVAELSRMCMVQNLAQYIFLHQICILDLNKGSGNQPICFVNYNSAL 2269
Db 2220 QHINHDVFDIYGLVAELSRMCMVQNLAQYIFLHQICILDLNKGSGHQPVCVNYNSTL 2279
QY 2270 QKMDSLDAMEGDVELEWEETM 2291
Db 2280 QKMDSLDAMEGDVELEWEETM 2301

RESULT 3
US-08-036-210-22
; Sequence 22, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mierock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 898 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-036-210-22

Query Match 37.8%; Score 4535.5; DB 1; Length 898;
Best Local Similarity 97.1%; Pred. No. 2.3e-269;
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYQNYKITTLRAQCKEWESECEVEYQKI 1488
Db 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYQNYKITTLRAQCKEWESECEVEYQKI 72
QY 1489 QVLYEAHLTEETVYGLKFRWRVRFQVAASNAGYNASNWI STKTLPGPPDGPENHVHV 1548

Db 73 QVLYEAHLTEETVYGLKFRWRVRFQVAASNAGYNASNWI STKTLPGPPDGPENHVHV 132
QY 1549 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKSEENKTIIEIKOLEIFTRY 1608
Db 133 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKSEENKTIIEIKOLEIFTRY 192
QY 1609 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQLTFLPPS 1668
Db 193 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQLTFLPPS 252
QY 1669 QPNGNIQVYQALVYREDDPTAVQIHNLISIIQKNTFVJAMLEGLGGHTYNI SVYAVNSA 1728
Db 253 QPNGNIQVYQALVYREDDPTAVQIHNLISIIQKNTFVJAMLEGLGGHTYNI SVYAVNSA 312
QY 1729 GAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLVSTTTIIRMPICIYSDHGPVKV 1788
Db 313 GAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLVSTTTIIRMPICIYSDHGPVKV 372
QY 1789 QVLATETGAQHDGNTWKYDAYFNKARPVFTNEGFPNPPCTEGTKFSGNBEIYIIGADN 1848
Db 373 QVLVETGAQHDGNTWKYDAYFNKARPVFTNEGFPNPPCTEGTKFSGNBEIYIIGADN 432
QY 1849 ACMIPGNEDKICNGPLPKPKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIIL 1908
Db 433 ACMIPGNEDKICNGPLPKPKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIIL 492
QY 1909 SVTLCILSIIILLGTAIPAFARIQKQEGGTYSPQDAEIIDTKLKDQIITVADLELKD 1968
Db 493 SVTLCILSIIILLGTAIPAFARIQKQEGGTYSPQDAEIIDTKLKDQIITVADLELKD 552
QY 1969 RLTR-----PISKKSFLQHVVEELCTNNNLKFQEFSELPKF 2004
Db 553 RLTRYSSFFFRKEIFVQLLSYRKSIRPISKKSFLQHVVEELCTNNNLKFQEFSELPKF 612
QY 2005 LQDLSSTDADLPWNRKRNFPNIKPYNNNNRVKLIADASVPGSDYINASYISGYLCPNEF 2064
Db 613 LQDLSSTDADLPWNRKRNFPNIKPY-NNNRVKLIADASVPGSDYINASYISGYLCPNEF 671
QY 2065 IATQGPLGTVGDFWRMWRTRAKTLVMLTQCFEGRIRCHOYWPEDNKPVTYVGDIVIT 2124
Db 672 IATQGPLGTVGDFWRMWRTRAKTLVMLTQCFEGRIRCHOYWPEDNKPVTYVGDIVIT 731
QY 2125 KLMEVDQIDWTIRDLKIERHGDQMTVROCNTFAWEHGVPENSAPLIHFVKLVRASRAHD 2184
Db 732 KLMEVDQIDWTIRDLKIERHGDQMTVROCNTFAWEHGVPENSAPLIHFVKLVRASRAHD 791
QY 2185 TTPMIVHCSAGVGRGVFIADHLTQHINDHDFVDIYGLVAELSRMCMVQNLAQYIPL 2244
Db 792 TTPMIVHCSAGVGRGVFIADHLTQHINDHDFVDIYGLVAELSRMCMVQNLAQYIPL 851
QY 2245 HQCILDLLSNKGSNQPICFVNYNSALQKMDSLDAMEGDVELEWEETM 2291
Db 852 HQCILDLLSNKGSNQPICFVNYNSALQKMDSLDAMEGDVELEWEETM 898

RESULT 4
US-08-449-609-22
; Sequence 22, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,609
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,210
FILING DATE: 23-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 898 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-449-609-22

Query Match 37.8%; Score 4535.5; DB 2; Length 898;
Best Local Similarity 97.1%; Pred. No. 2.3e-269;
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTQLRAQCKEWESECEVYQKI 1488
DB 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTQLRAQCKEWESECEVYQKI 72
QY 1489 QLYEAHLTEETVYGLKKFRWYRFQVAASTNAGYGNASNWIKTLPDPPDPPENHVHV 1548
DB 73 QLYEAHLTEETVYGLKKFRWYRFQVAASTNAGYGNASNWIKTLPDPPDPPENHVHV 132
QY 1549 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 1608
DB 133 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 192
QY 1609 SVVITAFGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQTLPLPS 1668
DB 193 SVVITAFGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQTLPLPS 252
QY 1669 QPNGNIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLEGLKGGHTYNI SVYAVNSA 1728
DB 253 QPNGNIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLEGLKGGHTYNI SVYAVNSA 312
QY 1729 GAGPKVPMRITWMDIKAPAPKTKPTPIYDATGKLLVTSTTIIRMPICYYSDDHGPIKNV 1788
DB 313 GAGPKVPMRITWMDIKAPAPKTKPTPIYDATGKLLVTSTTIIRMPICYYSDDHGPIKNV 372
QY 1789 QVLATGTAGHDGNTVYKVDAYFNKARPYFTNEGPNPCTGKTKFSGNEBIYIIGADN 1848
DB 373 QVLATGTAGHDGNTVYKVDAYFNKARPYFTNEGPNPCTGKTKFSGNEBIYIIGADN 432
QY 1849 ACMPGNEDKICNGPLKPKQYLFKFRATNIMQFTDSYSDPVKTLGSLERTVEIIL 1908
DB 433 ACMPGNEDKICNGPLKPKQYLFKFRATNIMQFTDSYSDPVKTLGSLERTVEIIL 492
QY 1909 SVTLCILSIIILGTALFAPARIQKQEGTYSPODAEIIIDTKLQDLQITVADLELXDE 1968
DB 493 SVTLCILSIIILGTALFAPARIQKQEGTYSPODAEIIIDTKLQDLQITVADLELXDE 552
QY 1969 RLTR-----PISKKSFLQHVHEELCTNNNLKQFESELPKF 2004
DB 553 RLTRYSSFFRRKEIFVIQLLSYRKSIPISKKSFLQHVHEELCTNNNLKQFESELPKF 612

QY 2005 LQDLSSTADLPWNRKAKRFPNKKPYNNNNRVVKLIADASVPGSDYINASYISGILCPNEF 2064
DB 613 LQDLSSTADLPWNRKAKRFPNKKPYNNNNRVVKLIADASVPGSDYINASYISGILCPNEF 671
QY 2065 IATQGPLGCTGDFWRWVWETRAKTLVMLTQCFEKGRIRCHOYWPEDNKPVTTVFGDIVIT 2124
DB 672 IATQGPLGCTGDFWRWVWETRAKTLVMLTQCFEKGRIRCHOYWPEDNKPVTTVFGDIVIT 731
QY 2125 KLMEVDQIDWTIRDLKIERHGDCTVRCNFTAWPEHGVPENSAPLIHFVKLVRAHRAHD 2184
DB 732 KLMEVDQIDWTIRDLKIERHGDCTVRCNFTAWPEHGVPENSAPLIHFVKLVRAHRAHD 791
QY 2185 TTPMIVHCSAGVGRGTGVFIALDHLTQHIINDHDFVDIYGLVAELRSRMCWQNLQYIFL 2244
DB 792 TTPMIVHCSAGVGRGTGVFIALDHLTQHIINDHDFVDIYGLVAELRSRMCWQNLQYIFL 851
QY 2245 HQCIILDLISNKGNOPICFVNYSAQKMDSLDAMEGDVELEWEETM 2291
DB 852 HQCIILDLISNKGNOPICFVNYSAQKMDSLDAMEGDVELEWEETM 898

RESULT 5

US-09-361-096A-22
Sequence 22, Application US/09361096A
Patent No. 6492495
GENERAL INFORMATION:
APPLICANT: MOLLER, NIELS P.H.
APPLICANT: MOLLER, KARIN B.
APPLICANT: ULLRICH, AXEL
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
FILE REFERENCE: 038602/0686
CURRENT APPLICATION NUMBER: US/09/361,096A
CURRENT FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 08/449,609
PRIOR FILING DATE: 1995-05-24
PRIOR APPLICATION NUMBER: 08/036,210
PRIOR FILING DATE: 1995-03-23
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 898
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2
US-09-361-096A-22

Query Match 37.8%; Score 4535.5; DB 4; Length 898;
Best Local Similarity 97.1%; Pred. No. 2.3e-269;
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;
QY 1429 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTQLRAQCKEWESECEVYQKI 1488
DB 13 VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFONYKITTQLRAQCKEWESECEVYQKI 72
QY 1489 QLYEAHLTEETVYGLKKFRWYRFQVAASTNAGYGNASNWIKTLPDPPDPPENHVHV 1548
DB 73 QLYEAHLTEETVYGLKKFRWYRFQVAASTNAGYGNASNWIKTLPDPPDPPENHVHV 132
QY 1549 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 1608
DB 133 ATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTIIEIKDLEIFTRY 192
QY 1609 SVVITAFGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQTLPLPS 1668
DB 193 SVVITAFGNISAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKEQTLPLPS 252
QY 1669 QPNGNIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLEGLKGGHTYNI SVYAVNSA 1728
DB 253 QPNGNIQVQALVYREDDPTAVQIHNLSTIIQKNTFVIAMLEGLKGGHTYNI SVYAVNSA 312
QY 1729 GAGPKVPMRITWMDIKAPAPKTKPTPIYDATGKLLVTSTTIIRMPICYYSDDHGPIKNV 1788

```
Db 313 GAGPKVPMRITMDIKAPARKTKPTBIYDAGKLLVTSTIIRMPICVYSDHGHPIKV 372
QY 1789 QVLATETGAHQDGNVTKWDAFNKARPFTNEGFPNCTEGKTKFSNGNEIYIIGADN 1848
Db 373 QVLVTETGAHQDGNVTKWDAFNKARPFTNEGFPNCTEGKTKFSNGNEIYIIGADN 432
QY 1849 ACMIPOGNEKICNGPLKPKKQYLKFRATNIMGQFTDSYSDPVKTLGSLGSLERTVEIIL 1908
Db 433 ACMIPOGNEKICNGPLKPKKQYLKFRATNIMGQFTDSYSDPVKTLGSLGSLERTVEIIL 492
QY 1909 SVTLCILSIILLGTAIFAFARIRQKQEGTYSPODAEIIDTKLQDQITVADLELKD 1968
Db 493 SVTLCILSIILLGTAIFAFARIRQKQEGTYSPODAEIIDTKLQDQITVADLELKD 552
QY 1969 RLTR-----PISKSFLOHVEELCTNNLKFQEFSELPKF 2004
Db 553 RLTRYSSFFFRKEIFVIQLLSYRKSIFPISKSFLOHVEELCTNNLKFQEFSELPKF 612
QY 2005 LODLSSTDADLPWNAKRNFPNKPNNNRVVKLIADASVPGSDYINASYISGYLCPNEF 2064
Db 613 LODLSSTDADLPWNAKRNFPNKPYPY-NNNRVKLIADASVPGSDYINASYISGYLCPNEF 671
QY 2065 IATQGLPGTGVGDFWVWVWETRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVIT 2124
Db 672 IATQGLPGTGVGDFWVWVWETRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVIT 731
QY 2125 KLMEVDQIDWTIRDLKIERHGDQMTVRQCNFTAWPEHGVPENSAPLIHFVKLVASRAHD 2184
Db 732 KLMEVDQIDWTIRDLKIERHGDQMTVRQCNFTAWPEHGVPENSAPLIHFVKLVASRAHD 791
QY 2185 TTPMIVHCSAGVGRGTGFALDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFL 2244
Db 792 TTPMIVHCSAGVGRGTGFALDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFL 851
QY 2245 HQCILDLLSNKGSNPICFVNYSAQKQMSLDAMEGDVELEWEETM 2291
Db 852 HQCILDLLSNKGSNPICFVNYSAQKQMSLDAMEGDVELEWEETM 898
```

RESULT 6

```
US-08-036-210-15
; Sequence 15, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9050
```

```
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-036-210-15

Query Match 15.9%; Score 1903.5; DB 1; Length 402;
Best Local Similarity 93.5%; Pred. No. 1.2e-108;
Matches 362; Conservative 0; Mismatches 0; Indels 25; Gaps 2;

QY 1929 RIRQKQEGGTYSPODAEIIDTKLQDQITVADLELKDRLTR----- 1972
Db 17 RIRQKQEGGTYSPODAEIIDTKLQDQITVADLELKDRLTRYSSFFFRKEIFVIQL 76
QY 1973 -----PISKSFLOHVEELCTNNLKFQEFSELPKFLODLSSTDADLPWNAKRN 2024
Db 77 LSYRKSIFPISKSFLOHVEELCTNNLKFQEFSELPKFLODLSSTDADLPWNAKRN 136
QY 2025 PNKPNNNRVVKLIADASVPGSDYINASYISGYLCPNEFIATQGLPGTGVGDFWVW 2084
Db 137 PNKPYPY-NNNRVVKLIADASVPGSDYINASYISGYLCPNEFIATQGLPGTGVGDFWVW 195
QY 2085 TRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERH 2144
Db 196 TRAKTLVMLTQCFEKGRIHQYWPEDNKPVTVFGDIVITKLMEVDQIDWTIRDLKIERH 255
QY 2145 GDCMTVRQCNFTAWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGF 2204
Db 256 GDCMTVRQCNFTAWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGF 315
QY 2205 LDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPICFV 2264
Db 316 LDHLTOHINDHDFVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPICFV 375
QY 2265 NYSAQKQMSLDAMEGDVELEWEETM 2291
Db 376 NYSAQKQMSLDAMEGDVELEWEETM 402
```

RESULT 7

```
US-08-449-609-15
; Sequence 15, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
```

ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-449-609-15

Query Match 15.9%; Score 1903.5; DB 2; Length 402;
Best Local Similarity 93.5%; Pred. No. 1.2e-108;
Matches 362; Conservative 0; Mismatches 0; Indels 25; Gaps 2;
QY 1929 RIRQKQEGGTYSPQDAEIIIDTKLKDQITVADLELKDERLTR----- 1972
DB 17 RIRQKQEGGTYSPQDAEIIIDTKLKDQITVADLELKDERLTRSSFFRRKEIFVQL 76
QY 1973 -----PISKKSFLQHVVELCTNNNLKQEFSELPKFLQDLSSTADLPNRAKNRP 2024
DB 77 LSYRKSIIKPISSKKSFLQHVVELCTNNNLKQEFSELPKFLQDLSSTADLPNRAKNRP 136
QY 2025 PNIPYNNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWRMWE 2084
DB 137 PNIPY-NNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWRMWE 195
QY 2085 TRAKTLVMTQCEPKGRIRCHQWYWPEDNKPVTTFVGDIVITKLMEDVQIDWTIRDLKIERH 2144
DB 196 TRAKTLVMTQCEPKGRIRCHQWYWPEDNKPVTTFVGDIVITKLMEDVQIDWTIRDLKIERH 255
QY 2145 GDCMTVRQCNFTAPHEGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVT 2204
DB 256 GDCMTVRQCNFTAPHEGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVT 315
QY 2205 LDHLTQHINDHDPVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPICFV 2264
DB 316 LDHLTQHINDHDPVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPICFV 375
QY 2265 NYSALQKMSLDAMEGDEVELEWETT 2291
DB 376 NYSALQKMSLDAMEGDEVELEWETT 402

RESULT 8
US-09-361-096A-15
Sequence 15, Application US/09361096A
Patent No. 6492495
GENERAL INFORMATION:
APPLICANT: MOLLER, NIELS P.H.
APPLICANT: MOLLER, KARIN B.
APPLICANT: ULLRICH, AXEL
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
FILE REFERENCE: 038602/0686
CURRENT APPLICATION NUMBER: US/09/361,096A
CURRENT FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 08/449,609
PRIOR FILING DATE: 1995-05-24
PRIOR APPLICATION NUMBER: 08/036,210
PRIOR FILING DATE: 1995-03-23
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 401
TYPE: PRT
ORGANISM: Homo sapiens
US-09-361-096A-15

Query Match 15.9%; Score 1902.5; DB 4; Length 401;
Best Local Similarity 92.8%; Pred. No. 1.4e-108;
Matches 361; Conservative 2; Mismatches 1; Indels 25; Gaps 2;
QY 1927 FARIQKQEGGTYSPQDAEIIIDTKLKDQITVADLELKDERLTR----- 1972
DB 14 YSEIRQKQEGGTYSPQDAEIIIDTKLKDQITVADLELKDERLTRSSFFRRKEIFVI 73
QY 1973 -----PISKKSFLQHVVELCTNNNLKQEFSELPKFLQDLSSTADLPNRAKNRP 2022
DB 74 QLLSYRKSIIKPISSKKSFLQHVVELCTNNNLKQEFSELPKFLQDLSSTADLPNRAKNRP 133
QY 2023 RPNIPKPYNNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWRMWE 2082
DB 134 RPNIPKPY-NNNRVKKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWRMWE 192
QY 2083 WETRAKTLVMTQCEPKGRIRCHQWYWPEDNKPVTTFVGDIVITKLMEDVQIDWTIRDLKIE 2142
DB 193 WETRAKTLVMTQCEPKGRIRCHQWYWPEDNKPVTTFVGDIVITKLMEDVQIDWTIRDLKIE 252
QY 2143 RHGDCMTVRQCNFTAPHEGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVT 2202
DB 253 RHGDCMTVRQCNFTAPHEGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVT 312
QY 2203 IALDHLTQHINDHDPVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPIC 2262
DB 313 IALDHLTQHINDHDPVDIYGLVAELRSERCMQVONLAQYIFLHQCILDLLSNKGSNPIC 372
QY 2263 FVNSALQKMSLDAMEGDEVELEWETT 2291
DB 373 FVNSALQKMSLDAMEGDEVELEWETT 401

RESULT 9
US-08-036-210-11
Sequence 11, Application US/08036210
Patent No. 558233
GENERAL INFORMATION:
APPLICANT: Moller, Niels P.H.
APPLICANT: Moller, Karin B.
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
TITLE OF INVENTION: PHOSPHATASE
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSER: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/036,210
FILING DATE: 23-MAR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
STRANDEDNESS: single

```
;
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-036-210-11

Query Match 14.3%; Score 1713.5; DB 1; Length 322;
Best Local Similarity 99.7%; Pred. No. 3.7e-97;
Matches 319; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPN 2031
DB 4 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNFIATQGPLGTVGDFWRMVMWETRAKTLV 2091
DB 63 NNNRVKLIADASVPGSDYINASYISGYLCPNFIATQGPLGTVGDFWRMVMWETRAKTLV 122

QY 2092 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDCMTVR 2151
DB 123 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDCMTVR 182

QY 2152 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 2211
DB 183 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 242

QY 2212 INDHDFVDIYGLVAELSRMCMVQNLQAYIFLHQCILDLLSNKGSNQPCFVNYSAQOK 2271
DB 243 INDHDFVDIYGLVAELSRMCMVQNLQAYIFLHQCILDLLSNKGSNQPCFVNYSAQOK 302

QY 2272 MDSLDAMEGDVELEWETTM 2291
DB 303 MDSLDAMEGDVELEWETTM 322

RESULT 10
US-08-449-609-11
; Sequence 11, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 322 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-449-609-11

Query Match 14.3%; Score 1713.5; DB 2; Length 322;
Best Local Similarity 99.7%; Pred. No. 3.7e-97;
Matches 319; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPN 2031
DB 4 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNFIATQGPLGTVGDFWRMVMWETRAKTLV 2091
DB 63 NNNRVKLIADASVPGSDYINASYISGYLCPNFIATQGPLGTVGDFWRMVMWETRAKTLV 122

QY 2092 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDCMTVR 2151
DB 123 MLTQCFEKGRIHQWPNEDNKPVTVFQDIVITKLMEDVQIDWTIRDLKIERHGDCMTVR 182

QY 2152 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 2211
DB 183 QCNFTAMPHEGVGPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGTGVFIALDHLTOH 242

QY 2212 INDHDFVDIYGLVAELSRMCMVQNLQAYIFLHQCILDLLSNKGSNQPCFVNYSAQOK 2271
DB 243 INDHDFVDIYGLVAELSRMCMVQNLQAYIFLHQCILDLLSNKGSNQPCFVNYSAQOK 302

QY 2272 MDSLDAMEGDVELEWETTM 2291
DB 303 MDSLDAMEGDVELEWETTM 322

RESULT 11
US-09-361-096A-11
; Sequence 11, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0866
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31D
US-09-361-096A-11

Query Match 14.3%; Score 1713.5; DB 4; Length 322;
Best Local Similarity 99.7%; Pred. No. 3.7e-97;
Matches 319; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPN 2031
DB 4 RPISKSFLOHVEELCTNNLKFOEFSELPKFLQDLSSTDADLPWNRKNRFPNPKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNFIATQGPLGTVGDFWRMVMWETRAKTLV 2091
DB 63 NNNRVKLIADASVPGSDYINASYISGYLCPNFIATQGPLGTVGDFWRMVMWETRAKTLV 122
```

QY 2092 MLTQCFEKGIRCHQWPDNKPVTGDIVITKLMEDVQIDWTIRDLKIERHGDCTVR 2151
DB 123 MLTQCFEKGIRCHQWPDNKPVTGDIVITKLMEDVQIDWTIRDLKIERHGDCTVR 182
QY 2152 QCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVTGFIADHLTQH 2211
DB 183 QCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVTGFIADHLTQH 242
QY 2212 INDHDFVDIYGLVAELRSEKCMQVQLAQYIFLHQICILDLSSNKGSNQPCFVNYALQK 2271
DB 243 INDHDFVDIYGLVAELRSEKCMQVQLAQYIFLHQICILDLSSNKGSNQPCFVNYALQK 302
QY 2272 MDSLDAMEGDVELEWETTM 2291
DB 303 MDSLDAMEGDVELEWETTM 322

RESULT 12

US-09-361-096A-47
; Sequence 47, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31D
US-09-361-096A-47

Query Match 12.8%; Score 1538.5; DB 4; Length 289;
Best Local Similarity 99.7%; Pred. No. 1.7e-86;
Matches 286; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1972 RPISKKSFLQHVBEELCTNNLKFQBEFSELPKFLQDLSSTADLPNNRAKNRPNIKPYN 2031
DB 4 RPISKKSFLQHVBEELCTNNLKFQBEFSELPKFLQDLSSTADLPNNRAKNRPNIKPY- 62
QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMWTETRAKTLV 2091
DB 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMWTETRAKTLV 122
QY 2092 MLTQCFEKGIRCHQWPDNKPVTGDIVITKLMEDVQIDWTIRDLKIERHGDCTVR 2151
DB 123 MLTQCFEKGIRCHQWPDNKPVTGDIVITKLMEDVQIDWTIRDLKIERHGDCTVR 182
QY 2152 QCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVTGFIADHLTQH 2211
DB 183 QCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVTGFIADHLTQH 242
QY 2212 INDHDFVDIYGLVAELRSEKCMQVQLAQYIFLHQICILDLSSNKGSN 2258
DB 243 INDHDFVDIYGLVAELRSEKCMQVQLAQYIFLHQICILDLSSNKGSN 289

RESULT 13

US-08-036-210-7
; Sequence 7, Application US/08036210
; Patent No. 5585233

; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,972
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-036-210-7
Query Match 10.0%; Score 1194.5; DB 1; Length 334;
Best Local Similarity 99.5%; Pred. No. 2.4e-65;
Matches 220; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1972 RPISKKSFLQHVBEELCTNNLKFQBEFSELPKFLQDLSSTADLPNNRAKNRPNIKPYN 2031
DB 4 RPISKKSFLQHVBEELCTNNLKFQBEFSELPKFLQDLSSTADLPNNRAKNRPNIKPY- 62
QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMWTETRAKTLV 2091
DB 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMWTETRAKTLV 122
QY 2092 MLTQCFEKGIRCHQWPDNKPVTGDIVITKLMEDVQIDWTIRDLKIERHGDCTVR 2151
DB 123 MLTQCFEKGIRCHQWPDNKPVTGDIVITKLMEDVQIDWTIRDLKIERHGDCTVR 182
QY 2152 QCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVTGFIADHLTQH 2211
DB 183 QCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVTGFIADHLTQH 242
RESULT 14
US-08-449-609-7
; Sequence 7, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,609
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,210
FILING DATE: 23-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Lealie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 334 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-449-609-7

Query Match 10.0%; Score 1194.5; DB 2; Length 334;
Best Local Similarity 99.5%; Pred. No. 2.4e-65;
Matches 220; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFQEFSELPKFLQDLSSTADLPWNRKNRFPNPKPN 2031
Db 4 RPISKSFLOHVEELCTNNLKFQEFSELPKFLQDLSSTADLPWNRKNRFPNPKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMWWETRAKTLV 2091
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMWWETRAKTLV 122

QY 2092 MLTQCFEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151
Db 123 MLTQCFEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182

QY 2152 QCNFTAPPEHGVGPENSAPLIHFVKLVASRAHDTTPMIVHC 2192
Db 183 QCNFTAPPEHGVGPENSAPLIHFVKLVASRAHDTTPMIVHC 223

RESULT 15

US-09-361-096A-7
Sequence 7, Application US/09361096A
Patent No. 6492495
GENERAL INFORMATION:
APPLICANT: MOLLER, NIELS P.H.
APPLICANT: MOLLER, KARIN B.
APPLICANT: ULLRICH, AXEL
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
FILE REFERENCE: 038602/0686
CURRENT APPLICATION NUMBER: US/09/361,096A
CURRENT FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 08/449,609
PRIOR FILING DATE: 1995-05-24
PRIOR APPLICATION NUMBER: 08/036,210
PRIOR FILING DATE: 1995-03-23
NUMBER OF SEQ ID NOS: 53

SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 7
LENGTH: 334
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: PTP-S31C
US-09-361-096A-7

Query Match 10.0%; Score 1194.5; DB 4; Length 334;
Best Local Similarity 99.5%; Pred. No. 2.4e-65;
Matches 220; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1972 RPISKSFLOHVEELCTNNLKFQEFSELPKFLQDLSSTADLPWNRKNRFPNPKPN 2031
Db 4 RPISKSFLOHVEELCTNNLKFQEFSELPKFLQDLSSTADLPWNRKNRFPNPKPY- 62

QY 2032 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMWWETRAKTLV 2091
Db 63 NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMWWETRAKTLV 122

QY 2092 MLTQCFEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 2151
Db 123 MLTQCFEKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVR 182

QY 2152 QCNFTAPPEHGVGPENSAPLIHFVKLVASRAHDTTPMIVHC 2192
Db 183 QCNFTAPPEHGVGPENSAPLIHFVKLVASRAHDTTPMIVHC 223

Search completed: June 30, 2005, 15:06:10
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 30, 2005, 14:45:54 ; Search time 65 Seconds
(without alignments)
3391.271 Million cell updates/sec

Title: US-10-673-885-2

Perfect score: 11985
Sequence: 1 MDPLIFLLFLLTGTSTQVD.....MDSLDMEGDVELEWEEETM 2291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10536	87.9	2302	2 T14328	protein-tyrosine-p
2	1319	11.0	2029	2 TDFPLK	protein-tyrosine-p
3	1188	9.9	1912	2 A56178	protein-tyrosine-p
4	1184	9.9	1894	2 C54689	protein-tyrosine-p
5	1181	9.9	1898	2 S46216	leukocyte antigen-
6	1173	9.8	1691	1 D54689	protein-tyrosine-p
7	1155.5	9.6	1897	1 TDHULK	leukocyte antigen-
8	1150	9.6	1615	2 B49502	protein-tyrosine-p
9	1150	9.6	1767	2 A49502	protein-tyrosine-p
10	1140	9.5	1907	2 S50893	protein-tyrosine-p
11	1127	9.4	1557	2 D41214	protein-tyrosine-p
12	1127	9.4	1630	2 C41214	protein-tyrosine-p
13	1107	9.2	1997	1 S12050	protein-tyrosine-p
14	1079	9.0	1863	2 S46217	protein-tyrosine-p
15	1033	8.6	2051	2 T30938	receptor tyrosine
16	882.5	7.4	1290	2 A56493	leucocyte common a
17	863	7.2	1238	2 S68700	HPTP beta-like tyr
18	855	7.1	1367	2 T21913	hypothetical prote
19	841	7.0	1226	2 JC7503	protein-tyrosine-p
20	840	7.0	1499	2 I50212	protein-tyrosine-p
21	834	7.0	1188	1 A57064	protein-tyrosine-p
22	833	7.0	1501	2 I58148	protein-tyrosine-p
23	827.5	6.9	1187	1 A53661	protein-tyrosine-p
24	823	6.9	1262	1 B48758	protein-tyrosine-p
25	822	6.9	1216	2 S60613	protein-tyrosine-p
26	822	6.9	1337	1 I38670	protein-tyrosine-p
27	820	6.8	1496	1 A48758	protein-tyrosine-p
28	807.5	6.7	1585	2 T19121	probable protein-t
29	800	6.7	2222	2 T13924	sdh protein - frui

30	794	6.6	1118	1 A49724	protein-tyrosine-p
31	783	6.5	1437	2 T31093	probable protein-t
32	770	6.4	1231	2 S53089	protein-tyrosine-p
33	770	6.4	1825	2 C88400	protein H19M22.1 l
34	770	6.4	1825	2 T32828	hypothetical prote
35	747	6.2	26926	1 I38344	titin, cardiac mus
36	745	6.2	1711	1 A55148	protein-tyrosine-p
37	744.5	6.2	593	2 S17671	protein-tyrosine-p
38	740.5	6.2	1452	1 S17669	protein-tyrosine-p
39	737	6.1	405	2 S68250	protein-tyrosine-p
40	727.5	6.1	405	2 I49372	protein-tyrosine-p
41	727	6.1	1440	2 JC6312	protein-tyrosine-p
42	721	6.0	1457	1 A48066	protein-tyrosine-p
43	703.5	5.9	1452	1 S17670	protein-tyrosine-p
44	691.5	5.8	6805	2 S20901	titin - rabbit (fr
45	668	5.6	1436	2 JC5290	protein-tyrosine-p

ALIGNMENTS

RESULT 1

T14328

protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type, GMCl precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T14328
R;Wright, M.B.; Hugo, C.; Seifert, R.; Distech, C.M.; Bowen-Pope, D.F.
J. Biol. Chem. 273: 23929-23937, 1998
A;Title: Proliferating and migrating mesangial cells responding to injury express a nov
A;Reference number: Z17986; MUID:98395110; PMID:9727007
A;Accession: T14328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2302 <WRI>
A;Cross-references: UNIPROT:O88488; EMBL:AF063249; NID:g3300095; PID:g3300096; PIDN:AAC

F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-2302/Product: protein-tyrosine phosphatase receptor type, GMCl #status predicted <

Query Match	87.9%;	Score	10536;	DB	2;	Length	2302;
Best Local Similarity	85.9%;	Pred. No.	0;				
Matches	1977;	Conservative	162;	Mismatches	151;	Indels	12;
						Gaps	4;
QY	1	MDPLIFLLFLLTGTSTQVDVNVVPGTRYDITISSIS-TTYTSPVTRIVTPNWKPGPP	59				
DB	2	MDPHFSFLLLIGTSSQVDVSSFDGTGYDITLSSVSATYSSPSVSRILATNWKPGPP	61				
QY	60	VFLAGRVGSAGILLISWNTPPNPNGRILISYIVKYKEVCPWMTQVYTVQVRSKPDSLEVLTT	119				
DB	62	VFLAGRVGSAGILLISWNTPPNPNGRILISYIVKYKEVCPWMTQVYTVQVRSKPDSLEVLTT	121				
QY	120	NLNPQTYTIKVAEAENSAGIVFSDPFLFQTAEAPGKVNLTVAYNASAVKLIWYLP	179				
DB	122	NLNPQTYTIKVAEAENAGIVFSDPFLFQTAEAPGKVNLTVAYNASAVKLIWYLP	181				
QY	180	QPNQKITSPKISVGHARGSGIVVKDVSIRVEDILTGLKPECNENSESFLMSTASPSPTLGR	239				
DB	182	QPNQKITSPKISVGHARGSGIVVKDVSIRVEDILTGLKPECNENSESFLMSTASPSPTLGR	241				
QY	240	VTPPSRTTHSSSLITQNEISSVWKEPISFVWVTHLRPTTYTLPEVSAATTGAYIDSTIVR	299				
DB	242	VTPVTRTQSSSTAARSKISSVWKEPISFVWVTHLRPTTYTLPEVSAATTGAYIDSTIVR	301				
QY	300	TPESVPEGPQNCVTGNITGKFSILWDPTTVTGKFSYRVLYGPGSRIIDNSTKDLKF	359				
DB	302	TPESVPEGPQNCINGNVTGKAFISWDPTTVTGKFSYRVLYGPGSRIIDNSTKDLRF	361				
QY	360	AFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDPGAVFDQLAARVESTQVITWKK	419				

Db 362 AFTHLTPFTMYDVVAAETISAGVGPKNLSVFTPPDPVPGAVFDLQIARVEATEIRITWRK 421
QY 420 PRQNGINGNOYKVLVPTGIIILENTLLTGNNEYINDPMAPEIVNIPEPMVGLYEGSAE 479
Db 422 PRQNGIISQYRKVSVLSTGVVLENTLLTGQDESISNEMPEIMLMDPMLGFTYEGSGE 481
QY 480 MSSDLHSLATFYNSHPDKNFAPARNRAEDQTSVPVTRNQYITDIAEQLSVYRRLVFP 539
Db 482 MSSDLHSPASFVNSHPHNDFPASTRAEQSPVPTTRNQYMTDITAEQLSVYRRLVFP 541
QY 540 THMISVSAPFTINGEGPPTVLVSTRQOVPSIIKIINYKNISSSSILLYWDPPPEYNGKI 599
Db 542 TEHTISVSAPFTINGEGPPTVLVSTRQOVPSIIKIINYKNISSSSILLYWDPPPEYNGKI 601
QY 600 THYTIYAMELDTNRAFQITTIIONSLITGLKYYTKYKMEVAASTHGDGSSLSSENDIFVR 659
Db 602 THYTIYATELDTNRAFQMTIVONSFLITGLKYYTKYKMEVAASTHGDGSSLSSENDIFVR 661
QY 660 TSEDEPESPQDVEVDVTADEIRLKWSPPEKPNGIIIAEVLVYKNIDLYMKNSTTDDI 719
Db 662 TPEDEPESPQDVQVTGVSPSELRLKWSPEKPNGIIIAEVLVYQNADTLFVKNTSTDDI 721
QY 720 ILRNLAPHTLYNISVRSTRPFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELS 779
Db 722 IISDLKPYTLXNISIRSYTRLGHGNQSSLLSVRTSETVPDSAPENITYKNISSGEIELS 781
QY 780 FLPPSSPNGIISKYTYLXRSNGNEERTINTTSLQNIKVLKCYQYIIEVSASTLKRG 839
Db 782 FLPPSPNGIISKYTYLXRSNSHEARTINTTSLAQTIQGLKYYTHYVIEVSASTLKRG 841
QY 840 VRSAPISIIITEADAPSPQDPSVKQLSGVTVKLSWQPPLEPNGIILYTVVYVNRSSILK 899
Db 842 IASRPISIIITEADAPSPQDPSVKQLSGVTVKLSWQPPLEPNGIILYTVVYVNDKSSLR 901
QY 900 TINVTETSELSLDLYNVSAYVTASTRFGDKTGSNIISFQTPPEGAPSDPPKOVYAN 959
Db 902 AINATEASLVSLDLYNDYGACVTASTRFGDGNARSSIINFRTPEGEPSDPPNDVHYVN 961
QY 960 LSSSSIIILFWTPSKPENGIIQYYSVYRNTSGTFMGNFTLHELNDPDMVTSTIIDKLT 1019
Db 962 LSSSSIIILFWTPSKPENGIIQYYSVYQNTSGTFVQNFLLQVTKESDNTVSARIYRLA 1021
QY 1020 IFSYTYFWLTASTSVGNKSSDIEVYTDQDIPPEGFVGNLTYESISSTAINVSWVPPAQ 1079
Db 1022 IFSYTYFWLTASTSVGNKSSDIIHYVTDQDIPPEGVGNLTYESISSTAIHVSWEPPSQ 1081
QY 1080 PNGLVYYSVLILQOT-PRHVRPPLVYERSIYFDNLEKYTDYILKIPSTEKGFSDTYT 1138
Db 1082 PNGLVFYLSLNLQOSPFRHMIPPLVYENSIDFDLEKYTDYIFKIPSTEKGFSETYT 1141
QY 1139 AOLYIKTEEDVPETSPIINTEKNLSSTVLLSWDPVPKNGALISYDLTLQGPENYSPT 1198
Db 1142 TQHLIKTEEDVPETPIINTEKNLSSTVLLSWDPVPKNGALIGVHLTLQGPANHFTV 1201
QY 1199 TSDNYIILBELSPFTLYSFFAAARTKGLGPSILFFYTTDESVPAPPQNLTLINCTSDP 1258
Db 1202 TSGNHVLEBELSPFTLYSFFAAARTKGLGPSILFFYTTDESAPLAPPQNLTLINYSDF 1261
QY 1259 VMLKWSPLPGGIVKIVYSPKIHHEHTDTIYKNISGFKTEAKVLGLEVPSTYSIRVSAP 1318
Db 1262 VMLTWSPLPGGIVKIVYSPKIHHEHTDTIVFYKNISGLQTDAKLEGPVSTYSVSASF 1321
QY 1319 TKVGNQNSNVKFTTQSSVPDVMQCMATSWQSVLKVKKDPPKANGIITQVMVYVE 1378
Db 1322 TKVGNQNSNVVEFTTQSSVPEAVRNICVARDWQSVRWDPPPKRTNGIILHYIMITYG 1381
QY 1379 RNSTKVSPQDMYTFIKLANTSYVFKVRASTVAGDGSSTCHVSTLPETVSPVPTNIAP 1438
Db 1382 GNSTKVSPDPTFTYFKLPTNSYVFEVRASTVAGGNSRCDISTLPETVSAPTNVAF 1441
QY 1439 SDVQSTSATLTWRPTILGYFQNYKITTLQRAQCKWESEBECVYKQIYLAHLTE 1498
Db 1442 SNVQSTSATLTWTKPTIFGYFQNYKITTLQRAQCKWESEBECIEHQDQVLYEANQTE 1501

QY 1499 ETVYGLKKFRWYRFOVAASTNAGYGNASNIISTKTLPGPPDPENHVHVATSPFSISIS 1558
Db 1502 ETVHGLKKFRWYRFOVAASTNVGYSNASEWISTQTLPGPDPENHVHVATSPFGINIS 1561
QY 1559 WSEPAVITGPTCYLIDVKSVDNDEFNISPIKNEENKTIIEIKDLEIFTRYSVVITAFQTN 1618
Db 1562 WSEPAVITGPTCYLIDVKSVDNDEFNISPLKNEENKTIIEINNLEVFTYSVIVITAFVGN 1621
QY 1619 ISAAVVEGSSSAEMLVTTLESAPKPPNNMTFKIPDEVTKFQLTFLPSPQNGNIQVYQ 1678
Db 1622 VSRAYTDGSSSAEVIITLLESVPKPPNNMTFKIPDEVTKFQLTFLPSPQNGNIRVYQ 1681
QY 1679 ALVYREDDPTAQIHNLSIIQKNTFVIAHLSGLKGGHYTINISVAVNSAGAGPKVPMRI 1738
Db 1682 ALVYREDDPTAQIHNFSIIQKTDTSIIAHLSGLKGGHYTINISVAVNSAGAGPKVQMRI 1741
QY 1739 TMDIKAPAPKTKPTPIYDATOKLIVTSTTIIRMPICYSDDHGPIKNOVLATETGAQ 1798
Db 1742 TMDIKAPAPKSKPIPIRDATGKLIVTSTTIIRMPICYNDHGHGPIRNVQVLVAETGAQ 1801
QY 1799 HOGNVTWKYDAYFNKARPYFTNEGPNPCTEGKTKFSGNEEYIIGADNACMIPGNEDK 1858
Db 1802 QDGNVTWKYDAYFNKARPYFTNEGPNPCCIIEGKTKFSGNEEYIIGADNACMIPGNEEK 1861
QY 1859 ICNGPLKPKKQYLKFRATNIMGQFTSDSYSDPVKTLGEGLSERTVEIILSVTLCLISII 1918
Db 1862 ICNGPLKPKKQYLKFRATNVGQFTSEYSDPIKTLGEGLSERTVEIILSVTLCLISII 1921
QY 1919 LLGTAIFAFAPARIRQKQEGGTYSPQDAEIIDTKLKLQDLITVADLELKDERLTR----- 1972
Db 1922 LLGTAIFAFAPARIRQKQEGGTYSPRAEIIDTKFKLDQDLITVADLELKDERLTRLLSYRK 1981
QY 1973 ---PISKSKFLOHVEELCTNNLKEESESSELPKELQDLSSTDAFLPWRANKRPPNIKP 2029
Db 1982 SIKPISKSKFLOHVEELCTNSNLKQEEFSELPKELQDLSSTDAFLPWRANKRPPNIKP 2041
QY 2030 YNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGLPGTVGDFWRVWVETRAKT 2089
Db 2042 Y-NNNRVKLIADVSLPGSDYINASYVSGYLCNFEFIATQGLPGTVGDFWRVWVETRTKT 2100
QY 2090 LVMLTQCFEKGIRCHQYWPEDNKPVTVFGDIVTKLMDVEDVQIDWTIRDLKIERHGDGMT 2149
Db 2101 LVMLTQCFEKGIRCHQYWPEDNKPVTVFGDIVTKLMEDIQIDWTIRDLKIERHGDGMT 2160
QY 2150 VQOCNFTAWPEHGVENSAPLIHFVKLVRAHADTTPMIHCSAGAGVGTGVFIADHLT 2209
Db 2161 VQOCNFTAWPEHGVENPTPLIHFVKLVRTSAHDTTPMVHCSAGAGVGTGVFIADHLT 2220
QY 2210 QHINHDHFDVYIYGLVAELASERMCMVQNLAQYIFLHQICILDLLSNKGSPICFVNYVAL 2269
Db 2221 QHINHDHFDVYIYGLVAELASERMCMVQNLAQYIFLHQICILDLLSNKGSGHQPCVFNYSTL 2280
QY 2270 QKMDSLDAMEGDEVELEWEETTM 2291
Db 2281 QKMDSLDAMEGDEVELEWEETTM 2302

RESULT 2

TDFFLK

protein-tyrosine-phosphatase (EC 3.1.3.48) DLAR precursor - fruit fly (*Drosophila melan-*
N;Alternate names: leukocyte antigen-related protein
C;Species: *Drosophila melanogaster*
C;Date: 14-Dec-1990 #sequence_revision 02-May-1994 #text_change 09-Jul-2004
C;Accession: A36182
R;Struelli, M.; Krueger, N.X.; Tsai, A.Y.M.; Saito, H.
Proc. Natl. Acad. Sci. U.S.A. 86, 8698-8702, 1989
A;Title: A family of receptor-linked protein tyrosine phosphatases in humans and *Drosop-*
A;Reference number: A36182; MUID:90046860; PMID:2554325
A;Accession: A36182
A;Molecule type: mRNA
A;Residues: 1-2029 <STR>
A;Cross-references: UNIPROT:P16621; GB:M27700; NID:g157811; PIDN:AAA28668.1; PID:g157811;

C:Genetics:

A;Gene: FlyBase:Lar

B;Cross-References: FlyBase:FBgn0000464

C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;

QY

C;Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane

F;132/Domain: signal sequence #status predicted <SIG>

F;132-2029/Product: leukocyte antigen-related protein #status predicted <MAT>

F;132-1377/Domain: intracellular #status predicted <EXT>

F;132-113/Domain: immunoglobulin homology <IMM1>

F;154-211/Domain: immunoglobulin homology <IMM2>

F;249-303/Domain: immunoglobulin homology <IMM3>

F;321-401/Domain: fibronectin type III repeat homology <FN3>

F;416-502/Domain: fibronectin type III repeat homology <FN3B>

F;514-599/Domain: fibronectin type III repeat homology <FN3C>

F;610-699/Domain: fibronectin type III repeat homology <FN3D>

F;708-802/Domain: fibronectin type III repeat homology <FN3E>

F;811-896/Domain: fibronectin type III repeat homology <FN3F>

F;909-993/Domain: fibronectin type III repeat homology <FN3G>

F;1006-1091/Domain: fibronectin type III repeat homology <FN3H>

F;1101-1198/Domain: fibronectin type III repeat homology <FN3I>

F;1378-1402/Domain: transmembrane #status predicted <TMM>

F;1403-2029/Domain: intracellular #status predicted <INT>

F;1417-2029/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F;1497-1718/Domain: protein-tyrosine-phosphatase homology <PTP1>

F;1786-2009/Domain: protein-tyrosine-phosphatase homology <PTP2>

F;57-111,161-209,256-301/Disulfide bonds: #status predicted

F;176/Binding site: carbonyl (Asn) (covalent) #status predicted

F;1670/Active site: Cys (phosphocysteine intermediate) #status predicted

F;1676/Binding site: substrate phosphate (Arg) #status predicted

F;1961/Active site: Cys (phosphocysteine intermediate) #status predicted

F;1967/Binding site: substrate phosphate (Arg) #status predicted

Query Match 11.0%; Score 1319; DB 1; Length 2029;

Best Local Similarity 24.3%; Pred. No. 3.7e-57;

Matches 530; Conservative 292; Mismatches 762; Indels 600; Gaps 85;

QY 270 VTHLRYTYVLEFVSAAATEAGYIDSTIV---RTPESVPEGPQNCVGTNITGKSFSLW 326

DB 5 MTAARPIAALSLLVLSLT---WTHPTIADAAHPPEIRKQNGVRGVASFYCAARG 61

QY 327 DPPTI-----VTGKSYRVELYGPSSG-----RILDNSTQDK 358

DB 62 DPPPSVWRKNGKVKSGTSRYTVLEQPGGISTRIEVPVGRAGDDAPYCEVAENGVDV 121

QY 359 FAFNTLTPTFTMDVYIAAETSAGTGPKSNISVPTPDVCAVFDLAEVSTQVRIWK 418

DB 122 SADATLT-----IYEGDKTPAG-----FPVIT---QGPET-----RVIEVGHT-VLMTCK 162

QY 419 KPRQNGIINQVRKVLVPETGIIILENTLLTGNNEYINDPMAPEIYVIVEPMVGLYEGSA 478

DB 163 AIGNPTP--NIYWK---NQTVDMSNPRYSKDGFL-----QIENSREEDQGYECVA 211

QY 479 EMSDDL-HSLATPIY---NSHPDKNPPANRAEDQTSVPTTRNQVITDIAEQLSYVR 534

DB 212 ENSMGTEHSKATLYYKVRVP---PTFSRPETISEVLMGSLNLSGCTIA-----VG 260

QY 535 RLVPFTHEMISVSFAITMGEGPTVLVSTRQOVPSIKIINVKNTSSSILLYWDPPY 594

DB 261 SPMPHVKKWKGSDLT-----PENEMPIGRNVQLINIQESA----- 297

QY 595 PNGKITHPTIYAMELTDNRAFIQTIDNSFLITGLKKYTKYKMRVAASHTDGESSLSEEN 654

DB 298 -----NYTC-----IAAST-----LGQID 311

QY 655 DIFVTSDEBPSSPODVEIDVTADEIRLKW---PPEKNGIIIAEYLYKNIDLYMK 712

DB 312 SVSVVKVQSLP-TAPTQVISEVTSVLEWSYKGPEDLQYVVIQYKPKNAQAFSEIS 370

QY 713 NTSITDIIILNLPHTLYNISVRSYTRFGHGNQVSLLSVRSYTSVTPSDAPENITKNS 772

DB 371 GIITMYVYVRLSPYTFEYFYVIANVNIIGRG--PPSAPATCTTGTETOMESAPRNVQRTLS 429

QY 773 SGEIELSFLPPSPNGIIEKKYIYILKRSNGNBERTINT---TSLTONIKVLKKYQYII 828

DB 430 SSTWITWEPPTENGQYGVYKYYTNSNQPEASWNSQMDNSLTTVSDVTFPHAIYTV 489

QY 829 EYASASTLKCEGVSAPISILITREDAFSPQPSVKQLSGVTVKLSSQWPLEPENGILY 888

DB 490 RVQAYTSMGAGMSTPVQKQGVF-SQPSNFRATDGETAVTLOWKTRTHSENIVHY 548

QY 889 TVYVW-----NRSLKTNVITETSLSDLDYVVEYSAYVTASTREFGDKGTGKNTIIISFOT 943

DB 549 ELY-WNDTVANQAHKRIKSNSE-AVTLGLYDPTLYIWLARSQRGEGAT-TPIPVRT 605

QY 944 PEGAPSDPPKDYVYANLSSSSILFWTPP--SKPNGIIQYISVYVY-----RNTSGTFMQN 996

DB 606 KQYVPGAPPRNTATATSTTSLSLWLPVPSRNGRIIYKVFVFEVGRDEDEATM-- 663

QY 997 FTILHELTNDPDMNTVTIIDKLTIFSYFTFWLTASTSVGNKNSDIIIEVYTDQDPIEGF 1056

DB 664 -TL-----NNT-SIVLDELKRWTEYKIWLAGTSVGDGPRSHPII-LRTQEDVP-GD 711

QY 1057 VGNLTYESISSTAINVSWVPPAQP--NGLVYVYVLSILOQTPRHVRPPLVYVE-----R 1108

DB 712 PQDVKATPLNSTSIHVSMKPPLEKORNGIIRGV-----HIHAQELRDSGKGLNE 761

QY 1109 SYTFD-----NLEKYDIYILKITPSTKSGSDTVAQIYIKTEEDVP----- 1150

DB 762 PFKFDVVDLTFNVITGLQDPTKYSIQVAALTRKGDGR-SAAITVKTGCPVVRPTVSLK 820

QY 1151 --ETSPILNTFKNLSTSVLLSWDPVPKNGALISYDL-----TLOGPNENYS 1196

DB 821 IMERBPV-----SIELEWERPAQTYGELRGYLRWGVKQDQALKEMLSGPO--- 867

QY 1197 FITSDNYIILBELSPFTLYSFFAAARTKGLGSPSSILFPYTDSEVSLAPPQNLTINCTS 1256

DB 868 -MTKRF-----DNLERGEVEYFVAGSNHIGIQETVKIIFQTECTGPGPSNIT----- 917

QY 1257 DFVWLKWSPLPGGIVKYVSPKIHHEHTDTIYKKNISGFKTEAKLVGLEPVSYSIRVS 1316

DB 918 ----- 917

QY 1317 AFTKVGNGQFNVNVKFTTQESVDPVQVQNMOCWATSWQSVLVKWDPPKK--ANGIITQYM 1374

DB 918 -----IRFOT-----PDV-----LCVTWDPPTREHNGIITRYD 946

QY 1375 VTV-----ERNSTKVSQDHWYTFIKLLANTSYVFKVRASTASAGEDESTCHVST 1424

DB 947 VQFHKKIDHGLSERNTL-----RKAVFTNLENTYIFRVRAYTKQAGSPSDKLIVE 1001

QY 1425 LPETVPSVPTNIAFSDVOSTSATLWIRPDTILGYFQNTYKI-TTQJRAQCKKEWEESECV 1483

DB 1002 TERDMGRAPMSLOAEATSRQTAIWM-EPVTSRGKLLGYKIFVTMTAVEDLDQWTK--- 1057

QY 1484 EYQKTOYLYEAHLTEET-VYGLKKPRWYFQVAASNAAGYGNASWISITKTLPGPDGPP 1542

DB 1058 -----TVGLTSESADLVNLEKFAQYAVAIAARPKNGLGRLESEKVTYRI---KPEDVP 1105

QY 1543 ENHVHVATSPFSGISWSBPVITGPTCYLIDVKS-----VDNDEFNISFIKSNEE----- 1593

DB 1106 LNLRAHDVSTHNTLSWSPPIRLT-PVNYKISFDAMKVPVDSQGFQSQVQPKREILKH 1164

QY 1594 -NKTIEIKOLEITFRYSVVITFTAGNI SAAYVEGKSSAEMI VTTLESAPKDPNNMTFQK 1652

DB 1165 VYKTHINELSPFTYVNVNSAI PSDYSY-----RPPTKI TTVTQMAAPQ-PMVKPDFY 1218

QY 1653 I--PREVTKFQITFLP-PSQPNGNTQVYQALVYREDDPTAVQIHNLSIIQKNTTFVIAML 1709

DB 1219 VYNGBEI-----LVILFQASEEYGPISHYLVVVPED-----KSNLH----- 1255

QY 1710 EGLKGHTYNI SVYAVNSAGAPKPMRITMDIKAPARPK-TKPTPIYDATGKLLVTSIT 1768

DB 1256 -----KIPDQFLTDLPLGRNKRPERNPAPYIA-AKFPORSIP 1291

QY 1769 ITRIMPICYSDDHGPIKNVQVLATETGAHQDGNVTWKYDAYFNKARPYFTNEGFPNPPC 1828

Db	1292	FTF-----HLGS-----GDYHN-----FTNR-----	1308
Qy	1829	TEGKTKFSGNEEYIIIGADNACMIQGVNEDKICNGPLKPKQKYLKFKRATNMQQ---FTD	1885
Db	1309	-----KLEREKRYRIFVRVAVDTFQKHLXTS	1334
Qy	1886	SDYS-----DPVKTLGEGLSERVEIILSVTLICILSIIIL	1920
Db	1335	SPSEFLSLDMREAPGERPHRDPNPAPEPEVSVNRKDE--PEILWVVLPLMWSTFIV	1392
Qy	1921	GTAFAFAIRKQKGGTYSPQDAEIIITKGLDQILITVADLEKDEBLTR-----	1972
Db	1393	STALIVLCVVKRRQPKCT--PDQAAVTRPLMAADLGAGFTSPDPVDMRLNFOPTPGMTS	1450
Qy	1973	--PISKSKFLOHVEELCTNNLKFQEFSELKFKLQDLSTADLPWNAKRNFPNKKY	2030
Db	1451	HPPIPISEFANHIERLKSNDKQFQYESIEBPQ--QOFTWDSNLEHNSKRYANVTAY	1509
Qy	2031	NNNRVKLTADASVPGSDYINASYISGYLCPNEFIATQGLPQTGDFWRMVWETRAKTL	2090
Db	1510	--DHSRVQLPAVEGVGSDYINANCYGRKHAYATQGLQETFDVDFWRMCWELKTAII	1568
Qy	2091	VMLTQCFEKGRIRCHOYVPEDNKPVTVFGDVIITKLMDVQIDWTIRDLKIERHG--DCM	2148
Db	1569	VMMTRLEERTRIKCDQWP--TRGTETYGIFVTITQELATYSIRTFQLCRQGFNDRR	1626
Qy	2149	TVRCQNFATWPEHGVDPENSAPLIHFVKLVASRAHDHDTMIVHCSAGVGRGTGVFALDHL	2208
Db	1627	EIKQLQETAWPDHGVDPHAPFLQFLRRCRALTTPPESGFVIVHCSAGVGRGTGCVIIVDSM	1686
Qy	2209	TOHNDHDFVDIYGLVAELSRMCMVQNLQAVIFLHQICILLNKSNGSQNPICFVN---	2265
Db	1687	LERMKHEKIIDYGHVTCRAQRNVMQTEDQIIFIHDAILEAI-----ICGVTEVP	1738
Qy	2266	----XSALQKM-----DSLDAME 2279	
Db	1739	ARNLHHLQKLILTEPGETISGME 1762	
RESULT 3			
A:56178			
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - human			
N:Alternate names: protein-tyrosine-phosphatase BTP-2			
C:Species: Homo sapiens (man)			
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004			
C:Accession: A56178; S12052; B44929			
R:Pulido, R.; Krueger, N.X.; Serra-Pages, C.; Saito, H.; Streuli, M.			
J. Biol. Chem. 270, 6722-6728, 1995			
A:Title: Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta isoforms.			
A:Reference number: A56178; MUID:95204468; PMID:7896816			
A:Accession: A56178			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-1912 <PUT>			
A:Cross-references: UNIPROT:P23458; GB:U38929; NID:G755652; PIDN:AAC41749.1; PID:G755653			
R:Krueger, N.X.; Streuli, M.; Saito, H.			
EMBO J. 9, 3241-3252, 1990			
A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatase			
A:Reference number: S12049; MUID:91006018; PMID:2170109			
A:Accession: S12052			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 390-1912 <KRU>			
A:Cross-references: GB:X54133; NID:G35789; PIDN:CAA38068.1; PID:G35790			
A:Note: the sequence from Fig. 5B is inconsistent with that from Fig. 5A in having 568-T			
R:Adachi, M.; Sekiya, M.; Arimura, Y.; Takekawa, M.; Itoh, F.; Hinoda, Y.; Imai, K.; Yac			
Cancer Res. 52, 737-740, 1992			
A:Title: Protein-tyrosine phosphatase expression in pre-B cell NALM-6.			
A:Reference number: A44929; MUID:92119637; PMID:1370651			
A:Accession: B44929			
A:Molecule type: mRNA			

A:Residues: 1756-1804, 'C', 1806-1845 <ADA>			
A:Cross-references: GB:S78086; NID:G243545; PIDN:AAB21147.1; PID:G243546			
A:Experimental source: pre-B cell NALM-6			
A:Note: sequence extracted from NCBI backbone (NCBIN:78086, NCBI:P:78087)			
A:Note: the authors did not report the entire codon for residue 90			
C:Genetics:			
A:Gene: GDB:PTPRD			
A:Cross-references: GDB:131384; OMIM:601598			
A:Map position: 9p24-9p24			
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; ogly			
C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane			
F:38-100/Domain: immunoglobulin homology <IMM1>			
F:140-209/Domain: immunoglobulin homology <IMM2>			
F:250-304/Domain: immunoglobulin homology <IMM3>			
F:711-811/Domain: fibronectin type III repeat homology <3PR>			
F:1293-1912/Domain: leukocyte common antigen cytosolic domain homology <LAC>			
F:1669-1992/Domain: protein-tyrosine-phosphatase homology <PTP>			
F:1553/Active site: Cys (phosphocysteine intermediate) #status predicted			
F:1553/Binding site: substrate phosphate (Arg) #status predicted			
F:1844/Active site: Cys (phosphocysteine intermediate) #status predicted			
F:1850/Binding site: substrate phosphate (Arg) #status predicted			
Query Match 9.9%; Score 1188; DB 2; Length 1912;			
Best Local Similarity 22.9%; Pred. No. 1.1e-50;			
Matches 487; Conservative 285; Mismatches 724; Indels 630; Gaps 76;			
Qy	296	TVTRPESVPEGP-PQNCVTGNITGKSFSLMDP-PTIVTGKFSYRV-----ELYGP	345
Db	18	TDATPPRPTRTFVDTQGVSGVASFICQATGDRPKIVNNKKGVSNQRFEVIEFDG	77
Qy	346	SGRIIDNSTKDLKFAFTNLTPF-----TMVDVYAAETSAGTGPKNISVFTPPDPVGA	399
Db	78	SSVL-----RIQPLTRDEAIYC-VASNNVGEISVSTRITVIREDOIPRG	124
Qy	400	--VFDL--QLAEVSTQV-----RITWKKPRQP-----NGIINQYRVKVLV	436
Db	125	PTIDMGPKLVKVERTATMLCAASGNDPBITWFKDPLVDTSNNNGRIKQLRSE---	181
Qy	437	PETGIILENTLITGNNEYINDPMAPEIVNVEPMVGLIEGSAEMSDLHSLA---TFIYN	493
Db	182	-----SIGTPIRG-----ALQIEQSESDQKCECVATNSAGTRYSPANLYVRE	227
Qy	494	SHPKNFPARNRAEDQTSVWVTRQYITDIAAEQLSVVIRLVPFTEHMISVSAPTIMG	553
Db	228	LREVRVPRFRSIPPTNHEIMPGGVNTICVA-----VGSFMPYKVMKLAEDLTPE	280
Qy	554	EGP--PTVLSVTRTQQ-----VPSSIKIINYKNI-----SSSSIL	586
Db	281	DMPIGHNVLELNDVRQSANYTCVMSTLGVIEAIAQITVKALPKPGTPVWTESTATSIT	340
Qy	587	LYWDP--PRYPNGKITHYTIYAMELDTNRAFOITID-----NSFLITGLKKYTKYKRV	640
Db	341	LTWDSGNPE-----PVSYYIIQHKPNSEELYK--EIDGVATTRYVAGLSYSDYEFVV	394
Qy	641	ASTHDGESSLSENDIFVRTSDEPSSPDQVEDVDVTADEIRLKWSPPEKNGIILAYE	700
Db	395	AVNNIGRGPPSE--PVLITQISEQAQSSAPRDVQARMLSSTTILVQWKEEENGLQGR	452
Qy	701	VLYKNIDTLVYKNTSTTDTI-----ILRLRPHLYNISVRSYTRFSGHGNQVSSLLSVRT	754
Db	453	VYVTWDPDPTQHVNNMKHNVDQSQITIGNLVQKTVSVKVLAFPTSIGDG-PLSSDIQVIT	511
Qy	755	SETVPDSPAENITYKNISSGEIEFLPSSPSSNGIILKTYIYLRKSNNGNEERTINTSLT	814
Db	512	QTGVV--GQPLNFKAPESETSLSSWTTPRS--DTIANYELVYKDGSHGEQRIITBPPTS	569
Qy	815	QNIKVLKKYQVYIIEVSATLKGEGVRSAPISILTEEDAPDPQDFSVKQLSGVTVKLS	874
Db	570	YRLQGLKPNLSYFFRLAARSPQGLGASTAEISARTMQSKPSAPPDQISCTSPSSSILVS	629
Qy	875	WQPPLEPNGLIILYYTVVWNRSLSKLTINVTTSTLSLSLDLDYNVVEYSAYVTASTRFGDKT	934

Db 630 WQPP----- 633
QY 935 GSNIIISQTEGAPDPKDVVYANLSSSIIILFWTPSPKNGIIQYISVYRTNTSGTFM 994
Db 634 -----PVEKQNGIIIEYSIKYTAVDG--- 654
QY 995 QNFTHLHNDPDMVTIILDKLTIFFSYTFWLTASTSVGNKGSDDIIEVTDQDIPR 1054
Db 655 EDDKPEHILG-IPSDTTKLLLEKEWTEYRITVTAHTDVGPGESLSVL-INTNEDVPS 712
QY 1055 GFVGNLTYSISSTAINVSN---VPPAQNGLVYVYVLILOQTPRHVRPPLV----- 1104
Db 713 GPRKVEVEAVNSTSVKWSRSPVKNQKQINGYQVHVVRMENGEPKQPMKQVMDLAD 772
QY 1105 -----TYERSIYFONLEKYTDYILKITPSTKSGFSDTYTQALYIKTEBEDVETSPIN 1157
Db 773 AQWEPDITTEHDMIIISGLQPEYSYSLTVTAYTTKGDGARSKPKL-VSTTGAVPKRLV- 830
QY 1158 TPKNLSTSVLLSWDPVPKPGNGAIISYDLTLQGNENYFITSNDVILIELSPFTLYSF 1217
Db 831 -INHTQMTALIQWHPVDVTFGLQYRLKF----- 860
QY 1218 FAAARTKGLGPSILFFYTDSEVPLAPQNLTLINCTSDFVWLKWSPLPGGIVKVYS 1277
Db 861 -----GRKMEPLTLEFSEK----- 877
QY 1278 FKIHETDTIYKNIISGFKTEAKLVGLBPVSTYSIRVSATKVGNGQFNSVVKFTQOE 1337
Db 878 -----DHFTATDIHGAS-----YVFLSARNKVGFGEM--VKEISIFE 915
QY 1338 SVP-DVVQNMCMATWSQSVLVKWDPPKKA--NGIITQY-----MVTVERNSTKV 1384
Db 916 EYPTGPPQLHSGGTTSTSVLSQSWQPPVLAERNGLITKTYLLYRDNINIPLEMEQ--LI 972
QY 1385 SPQDMHYTIKLLANTSVYFKVRASTAGSGDES-TCHVSTLPETPSV-PTNIAFSDVQ 1442
Db 973 VPADTTMTITGLKPDITTDVKNRAHSTKSGPGVSPVQRTLP--VDQPAKFNHVKAYM 1030
QY 1443 STSATLTWRPDTILGYFONYKIITQLRAQCKEWESEBCEVQKIQYLYEAHLTEETVY 1502
Db 1031 KTSVLLSWEP-----ENYNSAMPEFKIL-----YDDGKVEZ-----VDGRATQKLIV 1073
QY 1503 GLKFKFWYFQVAASNAG--YGNASNWISTKTLPGPPDGPENVHVATSPFSISWS 1560
Db 1074 NLKPEKSYFVL---TNRNSAGGLQHRVTAKTAP-----DVLRTKP----- 1112
QY 1561 EPAVITGPTCYLIDVKSVDNDEFNISFIKSNEKNTIEIKLEIFRYSVVIATFGNIS 1620
Db 1113 -----AFI----- 1115
QY 1621 AAYVEKSSAEMIVTLESAPKDPNNMTFQKIPDBVTKFQLTFLPPSPQNGNIQVYQAL 1680
Db 1116 -----GKTNLDMITV--QLPEVPAN-----ENIKGYIIVLPLKSRK-----F 1154
QY 1681 VYREDDPTAVQIHNL--SIQKNTNTFVIAMBLGKGGHTYNTISVAVNSAGAPKVPMRI 1738
Db 1155 IKPWESPDEMBELDELKEISRK-----RSIRYGREVELK- 1189
QY 1739 TMDIKAPARKTPTPIYDATGKLLVTSITITIRMPICYSDDHGPKVQVLATETGAQ 1798
Db 1190 -----PYIAHFVLPTEFTL-----GDD-----K 1209
QY 1799 HDGNVTWKYDAVFNKARPYFTNEGPNPCTEGTKFSGNEBIYIIGADNACMIPGNEK 1858
Db 1210 HVG-----GFTN-----KQLSQGEVFFVLA---VMEHAEK 1239
QY 1859 ICNGPLPKKQYLFKFRATNIMGQFTDSYDPVKTLG-----EGLSERIVTEIL 1908
Db 1240 M-----YATSPYSDPVVSMDLDPQITDEESGLI-----WVV 1271
QY 1909 SVTLCLISLILGTAIFAFARIRQKQEGGTVSPQDAIID-----TKLKLQILITVADLE 1964
Db 1272 GPVLAVVFIICIVIAILLKRAESDSRKSIPNNKIPTSHHPTDPVELREL-----NQ 1327

QY 1965 LKDERLTRPISKKSFLQHVVELCTNNNLKAFQBEFSELPKFLQDLSTADLPMNRAKNRP 2024
Db 1328 TPGMASHPPIPILELADHIERLKANDNLKFSQYESIDPG-QQFTWEHNSLVNKPKNRY 1386
QY 2025 PNIKPVNNNRVKLTADASVPGSDYINASYISGLYCPNEFIATQGLPGTVGDFPMWME 2084
Db 1387 ANVIAY-DHSRVLSSAIEGIPGSDYVNYIIDYRKQNAIYIATQGSPLPETFGDFWRMIWE 1445
QY 2085 TRAKTLVMLTQCFEGRIRCHOYWPEDNKPVTVPFGDIVITKLMEDVOI-DWTIRDLKIER 2143
Db 1446 QRSATVVMVTKLEERSRVKCDQWP--SRGTTHGLVQVT-LLDTVELATYCVRTFALYK 1502
QY 2144 HG--DQMTVRQCNFTAWPEHGVPEPNSAPLIHFVKLVASRAHDHTTPMIVHCSAGVGTGV 2201
Db 1503 NGSSSEKREVRQFOFTAWPDHGVPEHPTPLAFLRRVKTNCNPPDAGPMVHCSAGVGTGC 1562
QY 2202 FIALDHLTQIHNDHPVDIYGLVABLSEKCMQVQLAQYIFLHQILDLNKNKSNQPI 2261
Db 1563 FVIDAMLEIRIKHEKTVDIYGHVTLMRAQNYMVQTDQYIFTHDALLEAVTCGNTEVPA 1622
QY 2262 CFVNYSAQKMDSLDAMEGDVLEWE 2287
Db 1623 RNL-YAIQKLTQIETGENVTGHELE 1647

RESULT 4

C54689

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form B precursor
N:Alternate names: MPTP delta type B/C
N:Contains: protein tyrosine phosphatase, receptor type delta, splice form C
C:Species: Mus musculus (house mouse)
C:Date: 25-Apr-1995 #sequence_revision 19-May-1995 #text_change 09-Jul-2004
C:Accession: C54689; B54689
R:Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.
Mol. Cell. Biol. 13, 5513-5523, 1993
A:Title: MPTP delta, a putative murine homolog of HPTP delta, is expressed in specialized
A:Reference number: A54689; MUID:93360986; PMID:8355697
A:Accession: C54689

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1894 <Miz>
A:Cross-references: UNIPROT:Q64487
A:Experimental source: brain; splice form B
A:Note: sequence inconsistent with nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBI:137487)
A:Accession: B54689
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352, 'H', 354-535, 'S', 537-601, 1002-1894 <M12>
A:Experimental source: brain; splice form C
A:Note: sequence inconsistent with nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:136527, NCBI:136530)
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; ogy

C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester by
F:45-107/Domain: immunoglobulin homology <IMM1>
F:245-299/Domain: immunoglobulin homology <IMM2>
F:317-399/Domain: fibronectin type III repeat homology <FN3A>
F:1278-1894/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F:1652-1874/Domain: protein-tyrosine-phosphatase homology <PTP2>
F:1536/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1542/Binding site: substrate phosphate (Arg) #status predicted
F:1826/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1832/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.9%; Score 1184; DB 2; Length 1894;
Best Local Similarity 22.8%; Pred. No. 1.7e-50;
Matches 490; Conservative 289; Mismatches 712; Indels 662; Gaps 80;

QY 263 KEPISFVTHLRPYTYILFEVSAATEAGYIDSTIVRPESVPEPPQNCVGTGNTGKSF 322
Db 12 RPPLSLALT-----PFLCACAETPPRF-----TRTPVD-----QTGVSQGVASFIC 52

Qy	323	SILWDP--PTIVTGKFSVRV-----ELYGPSGRILDNSTKOLKFAFTNLTPF-----	367
Db	53	QATGDPRPKI VWNKGGKVSQNRFEVIEFDGSGSVL-----RIQPLRTPRDE 100	
Qy	368	TMVDVYIAAETISAGTGPKSNISVTPPDVPCA--VFDL--QLAEVESTOV-----	413
Db	101	AIYEC-VASNNVGEISVSTRILVRBEOQIPRGFTIDMGPOLKVVERTATWMLCAASGN 159	
Qy	414	--BITWKKPRQP-----NGIINQYRKVLVPETGIIILENTLLTGNNXYINDMPAEIV 464	
Db	160	PDPEITWFKDFLPVDTSNNGRIKQLR-----SESIG-----ALQIE 196	
Qy	465	NIVBPMVLGEGSABMSDLHSLA---TFIYNSHPDKNFPARNRAEDQTSVVTTRNOYI 521	
Db	197	QSEESDQKYECAVTSAGTRYSPANLYVRVRPFRPSIPPTNH-----EIMPGGSVNI 250	
Qy	522	TDIAAEQLSVYIRRLVPETEHIVSFAFTINGEP--PTVLVSVTRTQO-----	567
Db	251	TCVA-----VGSMPYVKNWLGAEULTPDEDDMPIGRNVLELNDVQSNAYTCVAMSTL 303	
Qy	568	-VPSIKIINYKNI-----SSSSILLYWDPPEYENGKITHYTIYAMELDTNRA 614	
Db	304	GVIEAIAQITVKALPKPGTPVWTESTATSITLWDSGN--PGPVSYIIL--QEKPNSE 359	
Qy	615	FQITIID-----NSFLITGLKKYTKYKWRVAASTHDGESSISEENDIFVRTSEDEPESPQ 670	
Db	360	EPYKEIDGIATTRYSVAGLSYSDYEFRRVAVNNIGRGPASE--PVLTOTSEQTSPSSAPR 417	
Qy	671	DVEVIDVTADRIELKWSPPKPNIGII IAYEVLVKNIDTLYMKNTSTTDI-----ILRNL 724	
Db	418	DVQARMLSSSTIILVQWKEPEEPNGOIQGYRYVTWDPTHQVNNMKGVADSOQTITLGNL 477	
Qy	725	RPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSFLPPS 784	
Db	478	VPQKTSYVKVLAFTSIGDG-FLSSDIQVITQGVPGOPLNFKAEPESETSILLSWTPPR 535	
Qy	785	SPNGILIKYIYLYKRSN--GNEERINTSLTONIKVLKKYTOYIIEVSASTLKEGVRSR 843	
Db	536	EDT--IASYELVDRDGOGEQRITIBPGTSYRLQGLKPNLSLYFRLSATSPQGLGASTA 593	
Qy	844	PISILTEEDAPDSPPPQSVKQLSGVTVKLSWQPPLEPNGII LYTVTVWNRSSLKTINV 903	
Db	594	EISARTWQ-KSAPQPDISCTSPSTSILVSWQPP-----	627
Qy	904	TETSLSLSDLDYNYVEYSAYVTASTRFGDKGTGSI IISFQTPEGAPSDPPKDVVYANLSS 963	
Db	628	-----	627
Qy	964	SIIILFWTPPPKNGIIQYYSVYVYRNTSGTFMONTLHETNDFDMNTVSIILDKLTIFS 1023	
Db	628	-----PVEKONGIITYSUKYAAVDG---EDYKPHEIIGNSSD-TTKYLLEOLEKWT 676	
Qy	1024	YTFWLITASTSVGNGKSDIIEVTDODIPGFGVGNLITYESISSTAINVSW---VPPAQ 1080	
Db	677	YRITVTAHTDVGPPELSVL-IRTEDEVSPGPRKVEAVEAVNATAVKVSRSPVPNKQH 735	
Qy	1081	NGLVYVYVSLIQOT--PRH--VRPLVYTERSYIFONLEKYTIYILKTIPTSEKGSPT 1136	
Db	736	QOIRGYQVHYVWMEGEPKSAMLKDVMLADAQDMIISGLQPETSYSLTIVATYTTKGDGAR 795	
Qy	1137	YTAOLYIKTEEDVETSPIINTFKNLSTSVLLSWDPPVKNGAIISVDLTLOQPNENYS 1196	
Db	796	SKPKL-VSTTGSVGPCKPLV--INHTQMNITALIQWHPPVDTFG-----LQYRLKFG 844	
Qy	1197	FITSDNYIILELSPFTLYSPFAARTKGLGSPSIIFFYTDSEVPPLAPPONLTLINCTS 1256	
Db	845	-----RKONEPLTLEFSEKE-----	860
Qy	1257	DFVWLKWSPPPLPGIIVKVSFKIHEHETDIYIYKNISGPKTEAKVGLPEPVSYSTRVS 1316	
Db	861	-----DHPTATDIIHKGAS-----YVFRLS 879	

RESULT 5
S46216
leukocyt
N;Altern

N;Contains: protein-tyrosine-phosphatase (EC 3.1.3.48)
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 20-Feb-1995 #sequence revision 20-Feb-1995 #text_change 09-Jul-2004
 C;Accession: S46216; S23252; A41032; A33154
 R;Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.
 Biochem. J. 302, 39-47, 1994
 A;Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-phosphatase
 A;Reference number: S46216; MUID:94347119; PMID:8068021
 A;Accession: S46216
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-1898 <ZHA>
 A;Cross-references: UNIPROT:Q64604; EMBL:L11586; NID:G205132; PIDN:AA37655.1; PID:G205132
 R;Hashimoto, N.; Zhang, W.R.; Goldstein, B.J.
 Biochem. J. 284, 569-576, 1992
 A;Title: Insulin receptor and epidermal growth factor receptor dephosphorylation by three
 A;Reference number: S23252; MUID:92287069; PMID:1599438
 A;Accession: S23252
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1361-1604; 1649-1898 <HAS>
 R;Pot, D.A.; Woodford, T.A.; Remboutsika, E.; Haun, R.S.; Dixon, J.E.
 J. Biol. Chem. 266, 19688-19696, 1991
 A;Title: Cloning, bacterial expression, purification, and characterization of the cytoplasmic
 A;Reference number: A41032; MUID:92011772; PMID:1918076
 A;Accession: A41032
 A;Molecule type: mRNA
 A;Residues: 1035-1072, 'S', 1074-1433, 'T', 1435-1638, 'N', 1640-1642, 'HT', 1645-1898 <POT>
 A;Cross-references: GB:M60103; NID:G205130; PIDN:AAA41510.1; PID:G205131
 R;Pot, D.A.; Woodford, T.A.; Remboutsika, E.; Haun, R.S.; Dixon, J.E.
 submitted to the Protein Sequence Database, December 1990
 A;Reference number: A33154
 A;Accession: A33154
 A;Molecule type: mRNA
 A;Residues: 1035-1072, 'S', 1074-1433, 'T', 1435-1638, 'N', 1640-1642, 'HT', 1645-1898 <PO2>
 C;Comment: Only the first of the two domains homologous with protein-tyrosine-phosphatase
 C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; tyrosine phosphorylation
 C;Keywords: duplication; glycoprotein; phosphoprotein; phosphoric monoester hydrolase; tyrosine phosphorylation
 F;1-27/Domain: (or 1-26) signal sequence #status predicted <SIG>
 F;28-1898/Product: (or 27-1898) leukocyte antigen-related protein #status predicted <MAT>
 F;28-1251/Domain: (or 27-1251) extracellular #status predicted <EXT>
 F;47-109/Domain: immunoglobulin homology <IMM1>
 F;149-209/Domain: immunoglobulin homology <IMM2>
 F;246-300/Domain: immunoglobulin homology <IMM3>
 F;318-400/Domain: fibronectin type III repeat homology <FN3A>
 F;413-499/Domain: fibronectin type III repeat homology <FN3B>
 F;511-593/Domain: fibronectin type III repeat homology <FN3C>
 F;606-695/Domain: fibronectin type III repeat homology <FN3D>
 F;708-799/Domain: fibronectin type III repeat homology <FN3E>
 F;811-895/Domain: fibronectin type III repeat homology <FN3F>
 F;906-990/Domain: fibronectin type III repeat homology <FN3G>
 F;1002-1079/Domain: fibronectin type III repeat homology <FN3H>
 F;1252-1275/Domain: (or 1259-1275) transmembrane #status predicted <TM>
 F;1276-1898/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 F;1366-1587/Domain: protein-tyrosine-phosphatase homology <PTP1>
 F;1655-1878/Domain: protein-tyrosine-phosphatase homology <PTP2>
 F;1454-207,156-207,253-298/Disulfide bonds: #status predicted
 F;117,250,295,721,957/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;1539/Active site: Cys (phosphocysteine intermediate) #status predicted
 F;1545/Binding site: substrate phosphate (Arg) #status predicted
 F;1830/Active site: Cys (phosphocysteine intermediate) #status predicted
 F;1836/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.9%; Score 1181; DB 2; Length 1898;
 Best Local Similarity 24.2%; Pred. No. 2.4e-50;
 Matches 478; Conservative 273; Mismatches 704; Indels 518; Gaps 72;
 QY 461 PEIVNVEPMVGLYEGSAEMSSDLHSLATFIYNHFDKPNPARRAEDQTSPPVT----- 515
 DB 33 PVFVKVPEDQIGLGG-----VAFVC-----QATGPKPRITWKKG 70

QY 516 --TRNQYITDIAAEQLSYVIRLVPF-----TEHMSVSAFTIMGE-GPPTVLSVTRTEQOV 568
 DB 71 KKVSSORFVEIPDDGAGSVLRIOPLRVORDEAIYECTATNSIAKLSVEEDQL 130
 QY 569 PS-----SIKIYNKNISSSILLYWDP-----PEYNGKITHYTIY 605
 DB 131 PSGFPTIDMGPKLVKVEKARTATMLCAAGNPDPPEISWPKDFLPVPASSNGRIKQLRSG 190
 QY 606 AMELDTNRAFIQTIDNSFLITGLKKYTKYKMAVAASHDGESSLSEENDIFVRTSDEP 665
 DB 191 ALQIESSEE-----SDQGYE-----CVATNSAGTRYSPANLYVRRVAP 232
 QY 666 ESS--PDQVEIDVTADEIRLKWSPKPNIGIIAYEVLKNIIDTLYMKNTSTDIILRN 723
 DB 233 RFSIPSSQEV-----PG-----NVNLTSCVAVGA 258
 QY 724 LRPHLTLYNLSVRSYTRFGH-----GNQVSSLLSVRTSETVPDSAPENITYKNIS----- 773
 DB 259 PMYVVMWMMGABELTKEDMPVGRNVLELSNVRS-----ANTCVAISSLGMEAT 310
 QY 774 GEIELSFLPPSSPNIGIKKYT-----IYLKRSNGNEE-----RTINT-----T 811
 DB 311 AQVTVKALPKPPIDLVVTTTATSVLTWDSGNTPEVSVFYGIQVRAAGTDGPFQEVGDVA 370
 QY 812 SLTONIKVLKKYQYIIEVSASTLKGEGVRSAPISILTEBDADSPQDFSVKQLSGTV 871
 DB 371 STRYSICGLSPSEYAFRLAVNSIGRGPPEAVRARTGEQAPSPRRVQARMLSASTM 430
 QY 872 KLSWQPLSPNGIILYTYV-----VNRSSIKTINVTETSI--ELSDLDYNYVEY 919
 DB 431 LVOMEPPPEBPNGLVRGYRYTYTTPDSRRPLSAMHKN-----TDAGLLITVGSLLPGITY 484
 QY 920 SAYVTASTRFGDKTGSNIISFOTPEGAPSDPKVYVYANLSSSIILFWTPSPKNGII 979
 DB 485 SLRVLAFTAVGDGP--SPITQVKTQGVPAQ--PADFQAKAESDTRQLSWLLPQER--II 541
 QY 980 QYVSVYRYMTSGTFMQNFTLHETNDNDMTVSTIIDKLTIFYTYFWLTASTSVGNK 1039
 DB 542 KYELVYV-----AADEGQGHKTFD---PTSSYTLLEDLKPDTLVHFLQAARSDLVGV--V 592
 QY 1040 SSDIIEVYTDQIDPEGFVGNLTYESISSTAINVSWYPPA--QNGLV-----FYVVSILIQ 1093
 DB 593 FTPTVEACTAQSTPSAPPQKVCVSTGSTRVSVWPPPADSRNGIITQVSVAYEADVGE 652
 QY 1094 QPVRHVRPLVTVRSIYFDNLEKYTDYILKIPTEKSGFDYTAQLYIKTBEDVPE 1153
 DB 653 DRKHVVVDGISEHSSWDLLGLEKTEYRWWRAHTDVG--PGSESPVLVTRTEDVP 711
 QY 1154 PIINTPKNISSTSVLSWDPPV--KNGAIISYDLTLOQGNENYSFITSNYIILELSP 1211
 DB 712 PRKVEVEPLNSTAVHVSWKLPVFNKQHGQIRGYVT-----YVRLNGEP 756
 QY 1212 FTLYSFAAARTKGLGPSILFFPYTDESVPPLAPQNLILNCTSFVWLKNSPSP 1271
 DB 757 -----RGQPII-----QDVMLAEAQ----- 771
 QY 1272 IVKVSFKIHEHTDIYKNIISGFKTEAKLGLPEVSTYIRVSAFTKVGNGNPNV 1331
 DB 772 -----ETTISGLTPTTYSITVAAYTKGDGAR--SKPK 803
 QY 1332 KFTQBSVPDVQNMCMATSMQSVLVKWDPPPKKANGIITQYMTYVER-----NSTKVS 1385
 DB 804 VVTTGAVPG--RPTMVVSTTAMHTALLQWHPKELPGLLGYRLQYRRADARPNTIDFG 862
 QY 1386 PQDHMTFTIKLANTSYVFKVRASTAGSDESTCHVSTLPETVPS--VPTNIAFSDVQST 1444
 DB 863 KDDQHFVTGLHKGATYIFRLAAKNRAGPGEFEKEITT--PEDAPSGFPQNLVTLG 921
 QY 1445 SATLTWIRDTIL-----GYFQNYKITTLQRAQCKEWESECEVEYKIQIYLYE----- 1493
 DB 922 TTLEAWDPP--VLAERNGRITNYTV-----YRDIINSQHEQLQNVGTD 961
 QY 1494 AHLTEETVYGLKFRWYRFPQAASTNAGYCNASNWISTKTLPGPDPGPPENVHVATSPF 1553

Db 962 VHL---TLGLKPTDYDIKVAHTSKGGLSPSIQSRTP-MEQVFAKNPRVAAMKT 1017
Qy 1554 SLSISWEP-----AVTGPCTVLIDVKSVDNDEFNIFSKNEENKTIIEIKDLEFTFY 1608
Db 1018 SVLLSWEVDSYKSAV---PPKILYNGQSVEDGHSMRKL-----IADLPQNTFY 1064
Qy 1609 SVVITAFGTGNISAAVVEGSSAEMIVTTLESAP-----KDPNNMTFQKT--PDEVTK 1659
Db 1065 SFVL--MNRGTSAGGLHLVSIKTAADLLPQKPLPASAFIEDGRFSLSPQVODPSLVRW 1122
Qy 1660 FQTLFLPPSPQNGNIQVQALVYREDPTAVQIHNLSIIQKTNFVIAMLEGLKGGHTYN 1719
Db 1123 FYIVVVPIDRVGNL-----LAPRWSTPEELEDEL-----LEAIEQEEQK 1164
Qy 1720 ISVAVNSAGAPKVPWMTMDIKAPARKTPTPIYDAGKLLVTSITTITIMPICYYS 1779
Db 1165 -----RRRRQAERLKPVAQAQVDELPTFTL----- 1191
Qy 1780 DGHGPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPFTNEGFPNPPCTEGTKFSGNE 1839
Db 1192 ---GDKKN-----YRGFYNR----- 1203
Qy 1840 EIYIIGADNACMPGNEDKICNGPLKPKQY-----LKFFRATNIMQFTDSYSDPVKTL 1895
Db 1204 -----PLSPDLSYQCFVLASKPEMDQKRYASSPYSDIIVQ 1240
Qy 1896 GEGLSERTVEIILSVT---LCILSIIILGTAFAPARIRQKQKGGTYSPQDAIITDKL 1952
Db 1241 VTPAQQQEPEMLWVTGPPVLAVILIIIVIALILLFKRK-----THSPSSKDEQSIGL 1293
Qy 1953 KLDOLITVADLELKDERLT-----RPISKKSFLQVHEELCTNNLKFQEEPSL-- 2001
Db 1294 K-DSSLAHSDPVEMLRNYQTPGMDHPPIPTDLDNIERLKANDGLKFSQYESIDP 1352
Qy 2002 -PKFLOQLSSTADLPMNRAKNRPNIKPYNNNRVKKLADASVPGSDYINASYISGYLC 2060
Db 1353 GQOFTWENSEV---NKPKNRYANVIAY-DHSRVLLTSIDGVFSGSDYINANYIDGYRK 1407
Qy 2061 PNEFIATQGLPOTGDFWRMVWETRAKTLMLTQCPEKGRIRCHQWYEDNKNPVTVFGD 2120
Db 1408 QNAYIATQGPLPOTMGDFWRMVWEQRIATVVMTRLEEKSRVKCDQYWPA--RGTYEYGL 1465
Qy 2121 IVITKLMEVDQI-DWTIRDLKIERHG--DCMTVROCNETAPWPHGVGPENSAPLIHFVKVL 2177
Db 1466 IQVT-LVDIVELATYTMRTFALHKSSESREURQFQWAPDHGVPEYPTPLAFLRV 1524
Qy 2178 RASRAHDTPPMIVHCSAGVGRGTGVFIADLHDLTQHINDHDFVDIYGLVABLRSERMCVQN 2237
Db 1525 KACNPLDAGPMVHCSAGVGRGTGCFIVIDAMLERMKHKTVDIYGHVTCRSQRNVMQT 1584
Qy 2238 LAQYIFLHCIDLLSNKSGNQPICFVNTSALQKM-----DSLDAWEGDVEL 2284
Db 1585 EDQYVFIHEALLE-AAMCGHTEVLARNLYAHIQKLGQVPPGSGSVTAMELEFKL 1636

RESULT 6
D54689
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form D precursor
N/Alternate names: MPTP delta type D
N/Contains: protein tyrosine phosphatase, receptor type delta, splice form A
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: D54689; A54689
R/Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.
Mol. Cell. Biol. 13, 5513-5523, 1993
A/Title: MPTP delta, a putative murine homolog of Hprt delta, is expressed in specialized
A/Reference number: A54689; MUID:93360986; PMID:8355697
A/Accession: D54689
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1691 <MI2>
A/Cross-references: UNIPROT:Q64487

A/Experimental source: brain
A/Note: sequence inconsistent with nucleotide translation
A/Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBIPI:136537)
A/Accession: A54689
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-398,799-1691 <MI2>
A/Experimental source: brain
A/Note: sequence inconsistent with nucleotide translation
A/Note: sequence extracted from NCBI backbone (NCBIN:136522, NCBIPI:136524)
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;
ogy
C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester h
F:42-95/Domain: immunoglobulin homology <IMM3>
F:114-196/Domain: fibronectin type III repeat homology <FN3A>
F:1479-1691/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F:1339/Active site: Cys (phosphotyrosine intermediate) #status predicted
F:1339/Binding site: Cys (phosphotyrosine intermediate) #status predicted
F:1623/Active site: Cys (phosphotyrosine intermediate) #status predicted
F:1623/Binding site: substrate phosphate (Arg) #status predicted
Query Match 9.8%; Score 1173; DB 1; Length 1691;
Best Local Similarity 26.4%; Pred. No. 5e-50;
Matches 426; Conservative 266; Mismatches 600; Indels 319; Gaps 72;
Qy 781 LPPSS-----PNCIIKKYTI-----YLRSGNEERT-----INTTSLTON-IKVLKK 822
Db 32 IPTTHNEIMPGGSVNITCVAGSPMPYKVMGLGAEDLTTPEDDMPIGRNVLELNDVRQSAN 91
Qy 823 YTOYIIIEVSASTLKGEGVRSAPISILTEDADPSPPQDPFSVKQLSGVTVKLSQWPPLEPN 882
Db 92 YTC-----VAMSTL---GVIEA-IAQITVKALP-KPGCTPVVTESTATSIITLWDSG-NP- 140
Qy 883 GIILYTVVVMNRSS---LKTIN-VTETSLESLDLYNVEYSAYVTASTRFGDGKGTGSMI 938
Db 141 GPVSVIIQHKPKNSEEPYKEIDGIATTRYSVAGLSYSDYFRVAVNVNIGRGPASEPV 200
Qy 939 ISFOTPEGAPSPDPKQVYVYANLSSSIIILFWTPPSKPGNGIIQYYSVYRNTSGTFMQNET 998
Db 201 LT-QTSEOTFPSSAPRDVQARMLSTTILVQWKEPEEPNGIQGYRYVYTMDFPTQHVNNWM 259
Qy 999 LHELTMDFNMVTVSTIIDKLTIFSVYTFMLTASTSVGNGKSDIIEVYTDQIDPEGFVG 1058
Db 260 KENVA-DSQITIGNLVPOKT---YSVKVLAFTSIGDGLSSD-IQVITQVGP-GQPL 312
Qy 1059 NUTYSISSTAINVSWVPPAQNGLVFYVSVILLOQTPRHPVPLVYR--RSIYFDNLE 1116
Db 313 NPKAPSESETSIILSWTPP-RSDTIASVE---LVDRDGDQGEQRITIBFGTSYRLOGLK 368
Qy 1117 KYTDYILKITPSTKGFSDTYTAQLYIKT-----EEDVPETSPIINTFKMLSSTVLL 1169
Db 369 PNLVYFRLSATSPQGLGAS-TAEISARTMQKPSAPPQDISCTSP-----SSTSLV 419
Qy 1170 SWDPP--VKPNGAIIISYDLTLQCPNENYSFISDNY-----IILEELSPFT 1213
Db 420 SWQPPPEVKONGIITEYSL-----KYAANDGEDYKPHIEICNSSDTTKYLLEQLEKWT 472
Qy 1214 LYSFPAARTRKGLGPSSI-LFFYTDSESVLPAQNLTLL--INCTSDFVWLKWSPPSLP- 1269
Db 473 EYRITVTAHTDVGWPPESLSVLIRTDDEVPSCGPRKVEEAVNATA--VKVSWR-SPEVN 529
Qy 1270 ---GGI--VKVYSPKHEHETDIYKNSIGPKTEAKLV-GLEPSTYSIRVSAPTKVN 1323
Db 530 KOHQGIRGQVHYVKNENGEPKSAAMLKQVMLADAQDMIIISGLOPETSLSLTAYTATKGD 569
Qy 1324 GNQFSNVVKTQESVPD---VVQNMOCMATSWOSVLVKWPPPKKANGIITQYMYTVR 1379
Db 590 GAR-SKPKLVSTGTVGPKRVLVNH-----TQMNTALIQWHPVDVTFG-LQGYRLKFR 642
Qy 1380 -----NSTKVSFQDHMYTFIKLLANTSVYFKVRASTASAGEDESTCHVSTLPTETVPS-V 1432
Db 643 KDMPELTLLEFSEKEDHFTATDIHKGASYVFRLSARNKVGFGSEMVKEIS-VPEIPTGF 701

Db 376 AFRVLAVNSIGRPPSEAVRAKTEQAPSPRRVQARMLASATMLVQVEPPEPPLVLR 435
QY 887 YTYVY-----VNRSSLLKINVTLSL--ELSDLDYXNVEYSAYYASTAFRGDGKT 934
Db 436 GYRVYTPDSRRPPNAWKHN-----TDAGLLTTVGSLPGTYSRLVLAFTAVGDGPP 489
QY 935 GNIISFQTEGAPSPKDVVYANLSSSS--ILFWTPPSKNGIIQYYSVYRNTSGTF 993
Db 490 -SPTIQVKTOQGVAPAD--FQAEVESDTRIQLSWLLPPQSR--IIMVELVYM-----AA 540
QY 994 MONFTLHETNPDNMTVSTIIDKLTFISYTFWLTASTSVGNKSSDIIIEYTDQDIP 1053
Db 541 EDEDQKHVTFD---PTSSYTLLEKPLDYRFQLAARSDMGV--VFTPTIEARTAQSTP 596
QY 1054 EGFVGNLTYESISSTAINSVWPPA--QPNGLVFYV---VSLILQOTPRHRVPLVTVYE 1107
Db 597 SAPPQKVMCSGTTTVRSWVPPPPADSRNGVTQYVAHEAVDGEDRGRHVVDGLSRH 656
QY 1108 RSIYFDNLEKYDYILIKITPSTEKGFSDTYTAQLYIKTEEDVPETSPIINTFNLSSTSV 1167
Db 657 SSWDLVGLKEWTEYRVVRAHTDVG--PGPESSPLVRLTDEDVPSGPPRKEVEPLNSTAV 715
QY 1168 LLSWDDPV--KNGAIIISDYLTLQGNENYFITSNYYILBELSFTLYSPFAAARTK 1225
Db 716 HVYWKLPVPSKQHGIRGQVT-----YVRLNGEP-----R 747
QY 1226 GLGPSSILFFYTDSEVPLAPPQNLTLNCTDFVWLKWSPLPGGIVKYSFKIHEHT 1285
Db 748 GL-----PII-----QDVLAAQ-----WRP-----ESEE 768
QY 1286 DRIYKNISGFTEAKLVGLEPVSTYSIRVSFTKVGNGQPSNVVKTFTQESVPDVQVQ 1345
Db 769 D-----YETISGLTETYSVVAATTKDQGR--SKPKIVTTTGNVPG--RPT 815
QY 1346 MQCMATSWOSLVKWDPPKKANGIIQYMTVER-----NSTKVPQDHMTYPIKILAN 1399
Db 816 MMISTAMNTALLQWHPKELPCELLGYRLQYCRADARPNTIDFGKDDQHFVTGLHG 875
QY 1400 TSVFVKRASTAGEDESTCHVSTLPETVPS--VPTNIAPSDVQSSATLTWRP----- 1453
Db 876 TTYIFRLAANRAGLEEKEIRT--PEDLPSPQNLHVTLTSTTELANDPPVLAER 934
QY 1454 -DTILGYFQNYK-ITTLQRAQCKEWESECEVEYQKIYLYEAHLTEETVYGLKPRWYR 1511
Db 935 NGRILSYTVFRDINSQ-----QELQNIITD--TRFTLTGLKPDPTD 975
QY 1512 FOVAASTNAGYNASNMISTKLPGPPDGPENNVHVATSPSISISWSEP-----AVIT 1566
Db 976 IKVRAWTSKSGPLSPSIQSRTMP--VEQVFAKNFRVAAAAMKTSVLLSWEVPDSYKSAV-- 1032
QY 1567 GPTCYLIDVKSVDNDEFNLSFKNSNEENTIEIKLEIFTRYSVVITATGHNISAAYVSG 1626
Db 1033 -PFKLYNGQSVEVDGHSRKL-----IADLPNTEYSFVLNM-----RG 1071
QY 1627 KSSAEM-IVTTLESAPKDPN-----NMTFOKIPDEVTKFOLTFLPPSQ 1669
Db 1072 SSAGGLQHLVSRINTADLLPHKPLPASAYIEDGRFPLSMHPHQDPSLVRFVIVVPIDR 1131
QY 1670 PNGNIQYQALVYREDPTAVQIHNLISIIQKNTVFVIAMLEGLKGHTYNI SVYAVNSAG 1729
Db 1132 VGGSM-----LTPRWSTPBELEDEL-----LEAIEQG----- 1159
QY 1730 AGPKVPMRITMDIKAPARKTKETPIYDATGKLLVTSTITITRMPICYSDDHGPIKNVQ 1789
Db 1160 -GSEQRR-----RRQAEALRKPVA--AQDLVLPETFTL-----GDKN-- 1195
QY 1790 VLATETGAQHDGNVTWKYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEYIIGADNA 1849
Db 1196 -----YGFYNR----- 1202
QY 1850 CMIPGNEDKICNGLPKPKQY-----LPKFRATNMQFTDSVSDPVKTLGEGLSRTVE 1905
Db 1203 -----FLSPDLISYQCFLVLSLKEPMDQKRYASSPSYDSBIVVQVTPAQOQEBP 1249

QY 1906 IILSVT---LCILSILLGTATFAFARIQKQKEGTSYSPQDAEIIDTKLKDQLITVAD 1962
Db 1250 EMLWTGVPVLVILIIIVIALILFKRK-----THSPSSKDEQSIGLK--DSLAAHS 1301
QY 1963 LBLKDBRLT-----RPISSKSLFQHVVELCTNNLKFQBESELSFKPKFQDLSSTDA 2013
Db 1302 DPVEVRRLLNYQTPGMRDHPPIITDLADNIERLKANDGLKFGQYESIDPG--QQFTWNS 1360
QY 2014 DLPWRKAKRPFNIKPYNNNRVKLIADASVPSGVYINASYISGLVCPNEFIATQGPLRG 2073
Db 1361 NLEVNKPKRYANVIAY--DHSRVILTSIDGVPESGVYINAYIDGRYKQWAYIATQGPLBE 1419
QY 2074 TVGDFWRMVWEVTRAKTLVMLTQCFEKGRIKCHQYVPEDNKPVTVFGDIVITKLMEVQI- 2132
Db 1420 TWGDFWRMVWEQRTATVVMTRLEKSRVKCQYWPA--RGYETCGLIQVT--LLDITVELA 1476
QY 2133 DWTIRDLKIERHG--DCMTVRCNFTAMPPEHGVENSAPLIHIFVKLVASRAHDTTPMIV 2190
Db 1477 TVTVRTFALHKSSEKRELRQFQWAMPDHGVPEYPTPILAFRRVKACNPLDAGPMV 1536
QY 2191 HCSAGVGTGVFIADHLTQHINDHDVVDIYGLVAELSERMCVQNLAQYIFLHOCIID 2250
Db 1537 HCSAGVGTGCFIVIDAMLERMKHKTVDIYGHVTCMSQRNMQVQTEQYVFIHEALLE 1596
QY 2251 LLSNKGSNOPICFVNYSALQKM-----DSLDAEGDVDEL 2284
Db 1597 AATCGHTEVPARNL-YAHIQKLGQVPPGESVTAMELEPKL 1635

RESULT 8

B49502
protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type 4E, splice form B precursor -
C:Species: Drosophila melanogaster
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: B49502
J. Biol. Chem. 268, 23964-23971, 1993
A:Title: Alternative splicing in a novel tyrosine phosphatase gene (DTP4E) of Drosophila.
A:Reference number: A49502; MUID:94043220; PMID:8226938
A:Accession: B49502
A:Molecule type: mRNA
A:Residues: 1-1615 <OON>
A:Cross-references: UNIPROT:Q9W4F5; GB:I20894
C:Genetics:
A:Gene: FlyBase:Ptp4E
A:Cross-references: FlyBase:Fgn0004368
A:Introns: 1605/3
C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III rep
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; recep
F:1254-1270/Domain: transmembrane #status predicted <TM>
F:1271-1615/Domain: intracellular #status predicted <INT>
F:1353-1573/Domain: protein-tyrosine-phosphatase homology <PTP>
F:1525/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1531/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.6%; Score 1150; DB 2; Length 1615;
Best Local Similarity 24.9%; Pred. No. 6.4e-49;
Matches 438; Conservative 266; Mismatches 611; Indels 442; Gaps 74;

QY 683 RLKAGPP-----EKPNGIITAYEVLKYNIDTLYMKNTSTTDIILNRPHTLYNISVRSYTR 739
Db 90 RIDYSPFPFPFNTTIPASDI-----GKDKFSRALPGTEYNFWL--YYIN 134
QY 740 FGHNQVSSLSVRSTETVPSAPENITYKNISGGEIELSFLPPSPN-----GIKKY 793
Db 135 STHQELTWVNIITTA---PD-PPANLSVQLRSKSAFITWRPPGSGRYSGFIRVLGLT 190
QY 794 TTYLKS---NGNEERTINTTSITQNIKV-LKKYQYIIEVSASTLKGSGVRSAPISILT 849
Db 191 DLPFRSISLEGETLQLSAKELTPGGSYQVQAYSVI-----QKSGSVATSRN 239
QY 850 BEDAPDSPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVYVWNRSSLKTNVTSTLE 909

Db 240 FTTKNTPGK-FIVWRNETTLVLMQPPF-PAGIYTHYRVSITPDATQSVLYVER--- 294
QY 910 LSDLNVNVEYSAYVASTAIFRCDGKTGNSIISFOTPEGAPSDPPKVVYYANLSSSSILFW 969
Db 295 -----EGEPPGPAQAAF----- 306
QY 970 TPSPKNGII--QYYSVYVRNTSGTFMQNFTLHETLNDPDMNVTVITIDKLTIFSYTF 1026
Db 307 -----KGLVPGREY-----NISQTVSED-ETSSVPTTARYLT----- 338
QY 1027 WLTAHSVGNKNSDIIIEVYTQDIPGEGVGNLTYES--ISSTAINVSVWPPAQNGLV 1084
Db 339 -----VPER-VLNVTFDEAYTSSFRVRWEPRTSYSEED 372
QY 1085 FYVVSILLOQTPRHVRPPLVYERSIYFONLEKYTDYILKITPSTKSGSDTYTAQLYIK 1144
Db 373 AYQV--MLSTSRRIFNVPAANGDSVYFD---YSIDL-----EPG---RTYEVVVK 415
QY 1145 TEEDVPTSP-----INTFKMLSTSVLLSWDPV--KPGCAIISDYLTLQ 1189
Db 416 TIADNVNVPASGEVTLPRPVASLGFLDRLDRNALHISWEPAETGRQDSYRISYEQTN 475
QY 1190 GPNENYFITSNDVYILEELSPFLYSFAAAR-----TRKGLGP-SSILFFYDES 1241
Db 476 ASEVPAPFVAESQITNTLTETLDSLLAGRYLIAVQALSGVASNASDITRYTRPA 535
QY 1242 PLAPPQNTLINCTDFWMLKWSPLPGGIVKVS-----FKIHEHETDIYKNSGFK 1297
Db 536 PLI--QELASID--QGLMLSWSD-----VNSRQDRYEVHYQNRGTREBTWATNE 582
QY 1298 TEAKLVGLBPVSTYSIRVSATFKVGNQFNSVVKTTQESVDVQVQNMOCWATSQSVL 1357
Db 583 TSLTIHYLPGSGYEVKVAH--ISHGVRSSEPHSYF--QAVFPKPPQNTLTQVHTNLVV 637
QY 1358 VKWDPPKANGIITQVMTVVERNSKVS-----QDHMYTFIKLLANTSYV 1403
Db 638 LHMWAPEGSD--FSEYV--RYRTDASPQRISGLHENEARKDMHY-----GERYL 685
QY 1404 FKVRASTAGDESTCHVSTLPTVPSVPTNIAPSDVOSTATLWIRPDTILGFQNY 1463
Db 686 VQV--NTVSFGVESP-HPELVNTWPPQVSVNVPLVDSRNLTLEWPRD---GHVDY 738
QY 1464 KITQLRAQCKEW--ESEECVEYQKIQVLYEAHLFEETV---YGLKFRWYRFOVAAS 1517
Db 739 TL-----KWPTEDEEDRVFKNVTOLED--LSSPSVRIPIEDLSPGQRYRFEVQAS 787
QY 1518 TRAGYGNASNWISTKLP-----GPPDGPENHVAVT-SP-----FS- 1554
Db 788 SN-GIRSGTTHLSTRMTPLIQSDVFIANAGHEQGDETITLSTYPTPADSTRFDIYRFSM 846
QY 1555 -----ISISWSEPAVITGPTCYLI-----DV 1575
Db 847 GDTIKDKEKLANDTERKLSFSLTGKLVNVTWTVSGVASLPVQVRVRLHPLPISDL 906
QY 1576 KSV-----DNDEFNISFIKSNEE-----NKTIIKOLEIFTRY 1609
Db 907 KAIQVAAREITLHTAPAGEYTDIFELQYLSADEAPQLLQNVTKNTEITLQGLRPVHNT 966
QY 1610 VVITATGNISAAVYVGSKAEMIVTL--ESAP-----KDPNNMTFKQIPDEVTK 1659
Db 967 FTWVVRSGSI-----QCTDPADVSVSTLMSAPISASYQTLTAPPGKVDYFQ-PSDVQP 1020
QY 1660 FQLTF--LPPSQPNQNIQVQALVYREDPTAVQIHNLSIIQKNTFTVIAMLEGKGGH 1716
Db 1021 GEVTFEWSLEPAEQHPIDYFRITCONADDAADVSYEPFV-----NATQGIKIDGLVPGN 1075
QY 1717 TYNISVAVNSAGAPKVPWRIITMDIKAPARKTKPTFYDATGKLVSTTITIRMPIC 1776
Db 1076 HYIFRQAKSALGYGAERHIOQWPIIAPVPEPSVTPL-----EVSRTSSIIETSFROG 1130
QY 1777 YSDDDHGPKNVQ-VLATETGAQHDG-NYTKWYD--AY-----FNKARYFTNEGFPNPP 1827

Db 1131 YFSNAHGMVRSYTIITIAEDVGKIASCLEMPMSQDVOAYTVWLPYQAIEPY-----NPF 1183
QY 1828 CT-EGTKPSGNEEYIIIGADNACMIPGNEDK-----ICNGPLKPKQYILFKFRATNIMQG 1882
Db 1184 LTSNGSRKSLAEHLITIGTANC-----DKHQAGYCNGLPFRAGTYRIKIRAFATDEDK 1236
QY 1883 FTDSDYSDVKTUGELSERTEIILSVTLCTILSITLLGTAIFAFARIQKQEGGTYSP 1942
Db 1237 FDTVYSSPITT-----ERSDTVIVAAT---VSALLVAMVLV-----VYCQ 1276
QY 1943 QDAEIIIDTKL-----DQLITVADLEKDBERLTRPISKSKSFLQHVBEELCTNNNLKFOEEF 1998
Db 1277 HRCOLIRRAKLARMQDELAALPEGYITP---NRPHVVKDFSEHYRIMSADSRSEEP 1333
QY 1999 SELPKFQLDSSTDADLPWNRKRPNNKPNKNNRVKLADASVPGSDIYINASYISG 2058
Db 1334 EELKHVGRDQACSFANLPCNRPKNRFTNLPY-DHSRFLQPVDDDDGSDYINANTMPGH 1392
QY 2059 LCPNEFIATQGPLPGTVGDFWBMVETRAKTLVMTQCEKGRIRCHOYWPEDNKPVTVF 2118
Db 1393 NSPREFIVTQGFPHSTREEFWRMCWESNRAIYMLTRCFEKGREKCDQYWPVD-RVAMFY 1451
QY 2119 GDIVITKLMEDVQIDWTIRDLKIERHGDCTVQRCNFTAWPERHGVGPENSAPLHFVKLVR 2178
Db 1452 GDIKVLQIIDTHYHDSISEFVWSRNCESRIMRHFHTTWDPGVEPPLSLVRFVRAFR 1511
QY 2179 ASRAHDTTPIVHCSAGVGTGVFIADHLTQHINDHDFVDIYGLVAELRSRMCVMQNL 2238
Db 1512 DVIQTDMPFIIVHCSAGVGRSGTFFIALDILQIHHSKDYVDIFGIVFAMRKEVFMVQTE 1571
QY 2239 AQVIFLHQICILLNSK 2255
Db 1572 QQYVCIHQCLLAVLECK 1588

RESULT 9

A49502
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 4E, splice form A precursor
C:Species: Drosophila melanogaster
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: A49502
R:Oon, S.H.; Hong, A.; Yang, X.; Chia, W.
J. Biol. Chem. 268, 23964-23971, 1993
A:Title: Alternative splicing in a novel tyrosine phosphatase gene (DTP4E) of Drosophila
A:Reference number: A49502; MUID:94043220; PMID:8226938
A:Accession: A49502
A:Molecule type: mRNA
A:Residues: 1-1767 <OON>
A:Cross-references: UNIPROT:Q9M4F5; UNIPROT:Q24495; GB:L20894
A>Note: authors translated the codon ATA for residue 1715 as Leu
C:Genetics:

A:Gene: ptp4E
A:Cross-references: FlyBase:FBgn0004368
C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III rep
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; recep
F:1254-1270/Domain: transmembrane #status predicted <TM>
F:1271-1767/Domain: intracellular #status predicted <INT>
F:1353-1573/Domain: protein-tyrosine-phosphatase homology <PTP>
F:1525/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1531/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.6%; Score 1150; DB 2; Length 1767;
Best Local Similarity 24.9%; Pred. No. 7.4e-49;
Matches 438; Conservative 266; Mismatches 611; Indels 442; Gaps 74;

QY 683 RLKMGSP---EKPNGLIIAYEVLVYKNIDFLYMKNTSTTDIILRNLPHTLYNISVRSYTR 739
Db 90 RIDYSPFPFGPEPNTIPASDI-----GKDIKFSRALPGTEYNFWL-YYTN 134
QY 740 FGHGQVSVLLSVRSETVPSAPENITYKNISGGEIHSFLPPSPN-----GIKKY 793
Db 135 STHQRLTWTVNITTA---PD-PPANLSVQLRSSKSAFITWRPPGSGRYSGFRIVLGLT 190

F:1554/Binding site: substrate phosphate (Arg) #status predicted
F:1839/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1845/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.5%; Score 1140; DB 2; Length 1907;
Best Local Similarity 25.5%; Pred. No. 2.6e-48;
Matches 424; Conservative 267; Mismatches 606; Indels 364; Gaps 66;

QY 750 LSVRTSETPDAPENITYKNISGIEBLSFLPPSPNGLIKKYTYLYKRSNGNEERT-- 807
DB 223 LYVRVRVAPRFSILPMSHEIMPGNNITCTVAVGSP-----MPYVMMQGAEDLTPE 275

QY 808 -----INTTSITONIKVLKXYQYIEVSASTLKGEVRSAPISILTEADAPDSPQDP 861
DB 276 DMPVGRNLELT-DVKDSANYTC-----VAMSSL-GVIEVAQITVYKSLPKAPGTP-- 325

QY 862 SVKQLSGVTVKLSQWPLENGIILYVTVVMNRS-----SLKTNVETSLSELSLDYN 916
DB 326 VVTENTATSTVTTWDSG-NPDP-VSYVYIEYKSKSQDGPYQIKE-DITTRYSIGLSFN 382

QY 917 VEYSAYVASTRFGDKTGSNIISFOTPEGAPDDPKDQVYVYANLSSSIIILFWTPPSKPN 976
DB 383 SEYEIWSVANSIGQPPSESVVT-RTGEQAPASAPRNQARMLSATTMLVQWEEBPVEN 441

QY 977 GIIQYYSVYRNTSGTGMQNFTHLTNDNDNTVSTIIDKLTIFSYTYFWLTASTSVGN 1036
DB 442 GLIRGRVYVTEPHEPVGVMQKHNV-DPSLITVGSLLDEET-----YTVRLAFTSVGD 496

QY 1037 GNKSSDIIEVYTDQDIPGEGVGNLTYESISTAINSVWVPVPAOPNGLVFVYVYLLIQPT 1096
DB 497 G-PLSDPIQVKTQGVV-GQPMMLRABAKSETSIGLSWAPROESVIKY---ELLFREGD 551

QY 1097 RHRVPLVYERSIYE-DNLEKYTDYILKITSTEGFSDDTYTAQYIKTEDEVPETSP 1154
DB 552 RG-REVGRTFDPTTAFVVEDLKENTYAFRLAARSPQGLG-AFTAVVHQRTLOAKSAPP 609

QY 1155 IINTFNKLASTSVLLSN--DPPVKNGAISYDLTLQ-----GNENYSFITSNYYI 1205
DB 610 QDVKCTSLASTALVSWRPPPTHTNGALVGSVYRPLGSEDDPKVENVNIPPTTQIL 669

QY 1206 LELSFTLYSFAAARTKGLP-SLILFFYTDSEVPLAPQNLTLINCTSDFWLWKS 1264
DB 670 LEALEKWTYRVTAVAYTEVGPSPSSVVRVDEDPVSPAPRKVEAEALNATKVLWR 729

QY 1265 PSLPLG---GIVKVSFKIHEHETDIYKNISGPKT-----EAKLVGLE 1306
DB 730 -SFTPGRQHGQIRGYO-----VHYVMEGAEGAPRINKDITMLADAQEWITNLQ 778

QY 1307 PVSTYSIRVSAFTKVGNGHQFNVVKTQESVPDVV-----QNMQCWATSQSVL 1357
DB 779 PETAYSITVAAVYTMKDG-----ARSKPKVVTKGAVLGRPTLSVQQTPEGSLL 827

QY 1358 VKWDPP-KKANGIIQYMTVVERNSTKVPQDHWYTFIKLLA-----NTSYVFK 1405
DB 828 ARWEPGDAEDPVLGYRLQFGR-----EDRAPATLEAAWERRFAAHPKAGATYVPR 880

QY 1406 VRASTAGSGDESTCHVSTLPTVPSVPTNI--AFSDVOSTSATLWIRPDIT--LGYFQ 1461
DB 881 LAARGGGLGEEAALIS-IPEDAPGFPQILGAGNVSAGSVLLRWLPVPAERNGAIL 939

QY 1462 NYKITTLQRAQKCKEWESECEVYQIQLYEAHLTEETVYGLKFRWYRFOVAASTNAG 1521
DB 940 KYTVSVREAGAPGPAFETELAAQAQ-----PGAETALTGLRLETGYELVRAHTRRG 993

QY 1522 YGNASNIWSTKTLPGPPDGP-----PENHVATSPFSISISWSEPAVITGPTC 1570
DB 994 -----PGFSPPLRYRLARDPVSPKRNFKVMKMTSVLLSWEFEDFNYSPTP 1040

QY 1571 YLIDVKSVDNDEFNISFIKSNEENKTIIE--IKDLEIFTRYVVVITAFGTNISAAY--VRS 1626
DB 1041 YKIQYNGLTLD-----VDGRYTKKLTITLKPHTFYNFVLTNRGSLGGLQQTVA 1090

QY 1627 KSSAEMIVTTLSEAPKDPNNMTFKIIPDEVTKFQTLFPLPPSPNGNIQVQALVYREDD 1686

DB 1091 RTAFNMLSKSPVAPK--PDN-----DGFIVVYLPDQO-----S 1122
QY 1687 PTAVQIHNLSIIQKTNFTFIAM-LBGLKGGHTYINISVAVNSAGAGPKVPMRITMDIKAP 1745
DB 1123 PVTVQ-----NYFIVMPLKSRG----- 1142
QY 1746 ARPQKPTPIYDATGKLLVTSTIIRMPICVYSDHGGPIKNVQVLATSTGAHQDGNVTK 1805
DB 1143 -----QFVLLGSPEDMDLEBLIQDISLQORH-----V 1171
QY 1806 WYDAFENKARPYFTNEGPNPCTEGTKFS-----GNEEYIITIGADNACMIPGNE 1857
DB 1172 RHRQLEVRPRPIA-----ARFSLPVAVFHNGKQY--GGFNRGLEPQH-- 1215
QY 1858 KICNGPLKPKQYLFKFRATNMGQ--FTDSYSDPV-----KTLGEGLSERTVEIIL 1908
DB 1216 -----RYVLVLAVLAKNEPTFAASFPDQPLDNDPQPIVDG--BEGLIWI 1262
QY 1909 SVTLCLISILILGTAFAPARIQKQEGGTVSPQDAEIIDTKLKLDQILITVADL---EL 1965
DB 1263 GPVLA VVFIICIVIAILLY-----KNK-----PDSKSKDSEPRTKCLLNADLAPHP 1310
QY 1966 KDERLTR-----PISKSKFLQHVLELCTNNNLKQEESESEL---PKFLQDLS 2009
DB 1311 KDPVEMRRINFOTPGMLSHPPITDMAHMERLXANDSLKSQYESIDPQOQQFTWHS 1370
QY 2010 STDADLPWNRKAKRPNFKPNNNRVKLIADASVPGSDYINASYSIGYLCNPEFTATOG 2069
DB 1371 NLEA-----NKPKNRYANVIAY-DHSRVILQPLEGMSDYINANYDVGRRQNAVATOG 1425
QY 2070 PLPGTVGDFMRWMTETRAKTLVMLTQCPEKGRIRCHQYWPEDNKPVTYVFGDIVITKMBD 2129
DB 1426 PLPETGDFMRWMBQORSATVVMVTRLEBSRIKCDQWY--NRGTETTYGFIQVT--LLDT 1482
QY 2130 VOI-DWTIRDLKIERHDC--MTVQCQNFETANPEHGVSPENSAPLHFKVLVRSRAHDTT 2186
DB 1483 MELATFCVTRTSLHNGSSGKREVRHFQTAWPDHGVPEYPTFLAFLRRVTCNPPDAG 1542
QY 2187 PMVHCSAGVGTGVFIADLHDLTOHNDHDFDIYGLVLAELSERMCMQNLQAYIFLHQ 2246
DB 1543 PIVHCSAGVGTGTCFVIDAMLERIKTEKTVDDVTVGHVTLMSQRNMYVQTEQYGFIEH 1602
QY 2247 CILDLLSNKSGNQPCFVNYSALQKMSLDAMEGDVELEWE 2287
DB 1603 ALLEAVSCGTEVPARSL-YTVIQKLAQVEPGEHVTGMELE 1642

RESULT 11

D41214

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 10D, short splice form precursor
C:Species: Drosophila melanogaster
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 09-Jul-2004
C:Accession: D41214; A41215
R:Yang, X.; Seow, K.T.; Bahri, S.M.; Oon, S.H.; Chia, W.
Cell 67, 661-673, 1991
A:Title: Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a sub
A:Reference number: A41214; MUID:92034988; PMID:1657401
A:Accession: D41214
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1557 <YAN>
A:Cross-references: UNIPROT:P35992; GB:M80465
R:Tian, S.S.; Tsoulfas, P.; Zinn, K.
Cell 67, 673-685, 1991
A:Title: Three receptor-linked protein-tyrosine phosphatases are selectively expressed
A:Reference number: A41215; MUID:92034989; PMID:1657402
A:Accession: A41215
A:Molecule type: mRNA
A:Residues: 1-904, 'L', 906-1125, 'Q', 1126-1165, 'YR', 1168-1171, 'A', 1173-1215, 'L', 1217-1456
A:Cross-references: GB:M80538; MID:g158644; PIDN:AAA28952.1; PID:g158645
C:Genetics:
A:Gene: FlyBase:Ptpl010

FEBS Lett. 282, 285-288, 1991
A:Title: Expression of a truncated protein-tyrosine phosphatase mRNA in human lung.
A:Reference number: S15818; MUID:91243813; PMID:1645282
A:Accession: S15818
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1872-1911, 'VHMLQK' <VR1>
A:Accession: S15819
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1872-1997 <VR2>
C:Genetics:
A:Gene: GDB:PTPRB; PTPB
A:Cross-references: GDB:127352; OMIM:176882
A:Map position: 12q15-12q21
C:Superfamily: protein-tyrosine-phosphatase, receptor type beta, fibronectin type III re
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; recept
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1997/Product: protein-tyrosine-phosphatase, receptor type beta #status predicted <M
F:23-1625/Domain: extracellular #status predicted <EXT>
F:1626-1642/Domain: transmembrane #status predicted <TM>
F:1643-1997/Domain: intracellular #status predicted <INT>
F:1727-1952/Domain: protein-tyrosine-phosphatase homology <PTP>
F:1904/Active site: Cys (phosphotyrosine intermediate) #status predicted
F:1910/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.2%; Score 1107; DB 1; Length 1997;
Best Local Similarity 23.8%; Pred. No. 1.2e-46;
Matches 519; Conservative 324; Mismatches 790; Indels 550; Gaps 101;

QY	373	YIAAEASAGT-----GPKSNISVFTPPDPCA-----VFIDLAEVEST 411
DB	29	FTLAESKASHSVSIQWRILGSPCNFLIYSSDTLGAALCFRTIDNTYGCNLDLQA- 87
QY	412	QVRITWKQPQNGIINQYRVKVLVPEGTGIIENLLTLLGNNEYINDMPAIVNIPVMV 471
DB	88	-----GTIYNFKIISLDEERTVWLQ-----TDLPPARFGVSK- 120
QY	472	GLYGEAEMSSDLHSLATIYNHSPDKNFPARNRAEDQTSVVTTFNQVITDIAEQLS- 530
DB	121	-----EKTSTGLHVMWT-----PSSGKVTSEYVQLFDENNOKIQGVQIQUEST 164
QY	531	---YVIRRLVPFTEHMISVSAFTIMGEGPTVLVSVRTR-QQVPSIIKIINYKNISSSIL 586
DB	165	WNEYTFNLTAGSKYNIATVVS-----GKRSESVYVNTGSTVSPVKDGI-STRANSL 219
QY	587	LYWDPPEYPNGKITHYTIYAMELDTNRAFQITID---NSFLITGLKKYTKYKMRVAAS 643
DB	220	ISW---SHSGNVVERYL--MLMDKGLVHGGVVDKHAITSYAFHGLSPGYLYNLTVMTEA 274
QY	644	HGESSLSBENDIFVTSDEPSSQDVEIDVTAD---EIRLKWSPPEKNGIIAY 699
DB	275	---AGLQYRWKLVRT-----APMEVSNLKVTDNGSLTSLKVKW---QRPPGNVDSY 320
QY	700	EVLYKNIDITLYMKNTS-----TTDILRLNRLPHTLYNLSVRSYTRFGHGNQVSSLL 750
DB	321	NI-----TLSHGKTIKESRVLAPWITETHFELVPEGLYQVTV-----SCVSGEL 365
QY	751	SVR---TSETVPDSPAENITYKNISSGEIELSPPLPSPNGIHKYTIYLKESNGNEERT 807
DB	366	SACKMAVGRTFPDKVANLEANNGMRSLVSVSPPA---GDWEQYRILL----- 412
QY	808	INTTSITONIKVLKKYTOYIIE-----VSASTLKG-----EGVRSAPISIL 848
DB	413	FNDSSVLLNITVKEETQYVMDPTGLVPGRQYVEVIVESGNLKNSERCQG-RTVPLAVL 471
QY	849	TEEDAPDPSQDFSVKQLSGVTVKLSWQPPPL---EPNGIILYVTVVNRSSL---KTIN 902
DB	472	-----QURVXKANTSLUSIMWQTPVAWEK-----YIIISLARDLLLIHKSUS 514
QY	903	VTETSLSLDLYNVEYSAYVTASTRFGDKGTGNSIISFTPTGAPSDPPKDVYANLS- 961
DB	515	KDAKEFTFTDLVPGKYMATVTSIS--GDLKNSSSVKGRVTPAQV-----TDLHVAQGM 567

QY	962	SSSIILFWTPPSKPNGLIIQIYISVYVRNTSGTFMQNFTLHETLNDNDNMTV--STIIDKLT 1019
DB	568	TSSLFTNWT---QAQGVFEYQV-----LLIHE-----NVVKNESISSETS 606
QY	1020	IFSYIY-----TFWLTAHSVGNKNSDIIEVYTDQDIPGEGFVGNLTIVYESISSTAINVSW 1074
DB	607	KYSPHSLKSGSLYSVVVTTVSGGISRRQV--VVEGRTPVSSVGVTVNNSGRNDVLSVSW 664
QY	1075	VPPAQPNGLVFTYYSL-----ILQOTPRHVRPPLVTVYERSIYFONLEKYTDYILKITST 1129
DB	665	L--VAPGDVDNYEVTLSHDKVQSL-----VIAKSVRECSFSLTTPGRLYTVTIT--T 714
QY	1130	EKGFSDTYTAQLYIKTERDVPETSPINFTFKNLSTSVLLSWDPDPVKNGGAIISYDLTLQ 1189
DB	715	RSKGVENHS-----FSQERTVDPKQGVSVNSARSNDYLRVSW---VHATGDFDHVEVIK 767
QY	1190	GNP---ENYSFTTSDNYIIEELSFPTLYSPFAAART-----RKLGFSSILFFYTDES 1240
DB	768	NKNNFIQTKSPKSENECVFQLVFGRLYSVTVTTKSGQYEANEQNG-----RT 817
QY	1241	VPLAPPQNLTLINCTSDFWLKWSPPLPGGIVKVSFKIHEHEHETDIYKKNISGFKTEA 1300
DB	818	IP-EPVKDLTLNRKSTEDLHVTVSGA---NGVDQYIEQLLFNDMKVFPFPHLVNTATEY 873
QY	1301	KLVLGPEPVSTYIRVSAFTKVGNGNQFNVVKFTTQESVPPDVQVQNMCMAT--SMQSVLVK 1359
DB	874	RPTSLTPGRQKILV--LITISDVQOQSAFIEGFT---VPSAVKNIHISPGNATDSLIVN 927
QY	1360	WDPPKKAAGIITQYVMTVERNSTKVSPQ---DHMY--TFIKLLANTSVYFKVRATSTAGE 1414
DB	928	WTP---GGGDVDSYTVSAFHSQKVDSTIPKHFVETPHRLEAGEQYQIMI--ASVSGSL 983
QY	1415	GDESTCHVSTLPETVSPVPTNIAESDVOSTSATLTWIRPDTILGVFYQVKI----- 1465
DB	984	KQINNVGRTVPASVQGVADNAYS---SYSLIVSW---QKAAGVAERYDILLTENGIL 1037
QY	1466	---TTLRAQCKEWESECEVEYQKIQYL-----YEALHTBET 1500
DB	1038	LRNTSEPAATYKQHKEDITPGKKYKIQLITVSGGLFSKEAQTEGRTVPAAVTDLRITENS 1097
QY	1501	VYGLKKFRW-----YRFQ-----VAAST 1518
DB	1098	TRHL--SFRWTASEGELSWNIFLYNPDGNLQERAQVDPLVQSFQNLQGRMYKQWIVT 1156
QY	1519	NAGYGNASNWISKTLPQPPD---GPPENVHVATSPESISISWSEPAVITGPTCYLIDV 1575
DB	1157	HSGELSNESFPGRTVPASVSHLRGSRN-----TTDSLWFFNWS--PA----- 1197
QY	1576	KSDVNDDEFNISPIKSEENK-TIEIKLEIF-----TRYSVVITAFITGNISAAAYVEG 1626
DB	1198	SGDFDTFELIILYNPGTKKENWKDKDTEWRFQGLVGRKYLWVTVHSGDLS-----N 1251
QY	1627	KSSAEMIVTTLESAPKPPNNMTFOKI-----PD---EVTKQLTFLPPS----- 1668
DB	1252	KVTAES-----RTAP--SPPLMSFADIANTSLAITWKPPDWTDYNDFELQWLPDALTV 1305
QY	1669	OPNGNIQVQALVY-----RE 1684
DB	1306	FNYPNRRKSEGRIVYGLRPGRSYQFNKTVSGDSWKTSKPIFGSVRTKPKDKIQLHCRP 1365
QY	1685	DDPTAV-----QIHNLISIIQKNTWTFVIAMLEGLKGGHTYNI 1720
DB	1366	QNSTAIACSWIPDSDDFGYSIECRKMDTQEVFSEKLEKESLLNIMM--LVPHKRYLV 1423
QY	1721	SVYAVNSAGAKVPMRITMDIKAPARKTKPTPIYDATGKLLVLTSTTITIRMPICYSD 1780
DB	1424	SI-KVQSAGMTSEVEDST--ITMDIRPPPPPHIRVNEKDVLSKSSINFVNCWSFSD 1480
QY	1781	DHGPIKNQVZLATETGA-----QHDGNVTKWTDAYFNKA--RPYTFNEGFPPPPCTE 1830
DB	1481	TNGAVKYFTVVVREADGSDDELKPEQOHP---LPSLEYRHRNASIRVYQTN--YFASKCAE 1535

1831 GKTKFSGNEEIIYIIGADNACWIPGNE- ---KICNGELPKKYOYLPKFA-TNI- ---MG 1881
1536 NPNNSNSKFNK-IGAEWE-SLGKRDPTQOKFCGDKPLKHTYAIRISIRATQLEDLK 1593
1882 QFTDSYSD- ---PVKTLGEGLSERTVEIILSVTLICILSIIILGTAIFAFARIQKQKE 1936
1594 EFTKPLYSDFPFLPITTESEPLF-GAIEGV-SAGLFLIGMLV- ---AVVALLICRQKVSH 1648
1937 GGTYSPODABIIDTKLKLQDLY-ADLELKDRLTR-PISKSKSFLOHVEELCTNNLKF 1994
1649 G- ---RERPSARLSIRDRPLSVHLNGLQGNKRTSCPIKINQFEGHFKLQADSNYLL 1703
1995 QERFSELKPKLQDLSSTADLPWNAKRNPPNKPNNNNRVKLIADASVPGSDYINASY 2054
1704 SKYTEELKQVGRQSCDIALPENRGNKRNINILPY-DATRVKLSNVDDPDCSDYINASY 1762
2055 ISGYLCNEPIATQGPLGTGVGFWRMVWETRAKTLVMLTQCFEKGRIHQWYEDNKP 2114
1763 IPGNPFREYIVTQGPLGTGDFWKMVWQNVHINVMVTCQVKEGRVKCDHYWADQDS 1822
2115 VTFPGDIVTKLMEVDIOWTIRDLKI- ---ERHGDCHVTRQCNFTAWPEHGVPENSAPLI 2171
1823 L-YGDLILQWLSVLPETIREFKICGEEQLDAHRLIRHPHYVWPDHGVETTSQSLI 1881
2172 HFVKLYR-ASRAHDTTPMIVHCSAGVGRGVFIALDHLTOHNDHDFVDIYGLVAELRS 2229
1882 QFVRTVRYDYNRSPGAGTIVHCSAGVGRGTGTGTRIALDRILQDLSKDSVDIYGAVHDLRL 1941
2230 ERMCMQNLQAQYIFLHQICLDLL 2252
1942 HRVHMVQTEQYVYLHCQVRDVL 1964

RESULT 14
S46217
protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat
N;Alternate names: leukocyte common antigen-related phosphatase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-May-1995 #sequence revision 03-Nov-1995 #text change 09-Jul-2004
C;Accession: S46217; S51174; A49104
R;Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.
Biochem. J. 302, 39-47, 1994
A;Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-phosphatase from rat
A;Reference number: S46216; MUID:94347119; PMID:8068021
A;Accession: S46217
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1863 <ZHA>
A;Cross-references: UNIPROT:Q64605; EMBL:L11587
R;Goldstein, B.J.
submitted to the EMBL Data Library, February 1993
A;Reference number: S51174
A;Accession: S51174
A;Molecule type: mRNA
A;Residues: 1-1788, 'G', 1790-1863 <GOI>
A;Cross-references: EMBL:L11587; MUID:9205134; PID:AAC37656.1; PID:9205135
R;Yan, H.; Grossman, A.; Wang, H.; D'Eustachio, P.; Mossie, K.; Musacchio, J.M.; Silverman, J. Biol. Chem. 268, 24880-24886, 1993
A;Title: A novel receptor tyrosine phosphatase-sigma that is highly expressed in the nervous system
A;Reference number: A49104; MUID:94043351; PMID:8227050
A;Accession: A49104
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-596, 'R', 598-603, 'I', 967-1788, 'G', 1790-1863 <YAN>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:139669)
C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; cyg
C;Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hydrolase
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1863/Product: protein-tyrosine-phosphatase #status predicted <MAT>
F;149-209/Domain: immunoglobulin homology <IMM1>
F;246-300/Domain: immunoglobulin homology <IMM2>

F;318-400/Domain: fibronectin type III repeat homology <FN3>
F;413-499/Domain: fibronectin type III repeat homology <FN3>
F;511-592/Domain: fibronectin type III repeat homology <FN3>
F;1244-1863/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F;1331-1552/Domain: protein-tyrosine-phosphatase homology <PTP>
F;1504/Active site: Cys (phosphotyrosine intermediate) #status predicted
F;1510/Binding site: substrate phosphate (Arg) #status predicted
F;1795/Active site: Cys (phosphotyrosine intermediate) #status predicted
F;1801/Binding site: substrate phosphate (Arg) #status predicted

Query Match 9.0%; Score 1079; DB 2; Length 1863;
Best Local Similarity 25.2%; Pred. No. 2, 7e-45;
Matches 416; Conservative 262; Mismatches 582; Indels 394; Gaps 66;

QY 750 LSVRTSETVDSAPENITYKNISSGIEISLFPSPNGIYKTYTYLKRSGNEERT-- 807
DB 223 LVVRVRVAPRPSILPMSHEIMPNGVNITCVAVSP-----MPYVKKMQGAEDLTPE 275
QY 808 -----INTSLTQNIKVLKTYTYIIEVSASTLKGEVRSAPISILTDEADPDSPPQDF 861
DB 276 DMPVGRNVLELT-DYKDSANYTC-----VAMSSL-GVIEAVAQITVKSLPKAPGTP--- 325
QY 862 SVKQLSGVTVKLSWQPLPENGILVYTVVWNRS-----SLKTINVTSLSLSLDYN 916
DB 326 VTENTATSIITWDSG-NPDP-VSYVIEYKSKSQDGPYQKE-DITTRISIGGLSPN 382
QY 917 VEISAYVTASTRFGDKTGSNIISFQTPGAPSDPPKQVYVYANLSSSSIIILFWTPPSKPN 976
DB 383 SEVEIWSAVNSIGQPPSPSVVT-RTGQAPASAPRNVQARMLSATTMIVQWEEFVEPN 441
QY 977 GLIQYVYVYRNTSGTFMONFTLHETNDPDMVTSTIIDKLTFISYTYFWLTASTSVGN 1036
DB 442 GLIRGYRVYTYMEPEHPVGNWQKHNV-DDSLLTTVGSLLEDET-----YTVRVLAFTSVGD 496
QY 1037 GNKSSDIEVYTDQDIPGFGVGNLTVESISSTAINVSWVPVPAQNGLVVYVYVLILOOTP 1096
DB 497 G-PLSDPIQVKTQGVV-GQPNVLRRAKASETSIGLSWAPQESVIKY---ELLFREGD 551
QY 1097 RHRVPLVTVYRSIYF--DNLEKYTDYILKITPTEKGFSDTYTAQYIKTBEDVETSP 1154
DB 552 RG-REVGRFTDPTAFVVEDLKPNTAYAPRLAARSQGLG-AFTAVVCQRTLQAKSPAP 609
QY 1155 INTFNLSSTSVLSSWDPVVPKNGAISDYDTLQGNENYFITSNDVNIIEELSPPFL 1214
DB 610 QDVKCTSLRSTAIL-----ILLEALEKWE 634
QY 1215 YSFFAAARTRKGLP-SSILPFYTVDSVPLAPPNTLINCTSDFVWLKWSPLPG--- 1270
DB 635 YRTAVAYTEVGPSPSSPVVTRDEDVPSAPPRKVEABALNATAIRVLR-SPTPGROH 693
QY 1271 GIVKVSFKIHEHETDIYKNGISGFKTEAK-----LVGLEPVSITYSI 1313
DB 694 GQIRGVQ-----VHYVMEG--TEARGPPRIKDIMLADAQEWITNLQPEATYSI 741
QY 1314 RVSAFTKVGNGQNFNVKFTQESVDDV-----QNMCMATSHQSVLVKWDPP- 1363
DB 742 TVAAVYTMKGDG-----ARSKPKVVTGKAVLGRPTLSVQQTPEGSLIARMEPPA 790
QY 1364 KKANGIITQYMTVVERNSTKVSQDHYTFIKLLA-----NTSVYVFKVRASTSA 1412
DB 791 DAAEDFVLGYRLQFGR-----EAAAPATLELAWERRFAAPAHKATYVRLAARGA 843
QY 1413 GEGDESTCHVSTLPETVPSVPTNI--AFSDVOSTSATLTWIRPDITLGFQNYKITTLQR 1470
DB 844 GLGEESAALS-IPEDAPRGFPQILGPAGNVSAGSVILRWLPVPAEGNGAIKYTVSR 902
QY 1471 AOKCKEWSBECVQKTYLYEAHLTBETVYGLKFKFRWYRFQVAASSTNAGVGNASNWS 1530
DB 903 ----BAGTPGPATETELAAAAQGAETALTQGLRPETAYELRVRAHTRG----- 949
QY 1531 TKTLPGPDGP-----PENVHVAVTSPFISISWSSEPAVITGPTCYLIDVKSVD 1579
DB 950 ----FGPFPPLRYRLARDPVSPKPKVIMKTSVLLSWEPFDNYSPTPYKIQTNGLT 1005

QY 1580 NDEFNIFIKSNEENKTIB--IKDLIFTRYSVVITAFGNISAAY--VEGSSAEIMVT 1635
Db 1006 LD-----VDGRITTKLTHLPHFTYFVLTNRGSSGLGLOOTVARTAFNLSG 1055
QY 1636 TLESAPKPPNNMTFOKIPDEVTKFQTLPLPSQPNNGNIQVYQALVYREDDDPTAVQIHL 1695
Db -1056 KPSVAFK--PDN-----DSSI-----VVYLPDGGSPVTVQY 1085
QY 1696 SIIQKNTTFVIAMLEGLKGGHTYNI SVYAVNSAGAPKVPMLRTMDIKAPARKTKPTPI 1755
Db 1086 FI-----VMVPLKSRGG----- 1098
QY 1756 YDATGKLLVTSITITIRMPICYYSDHGPKN-VQVLA--TETGAOHGDNVTKWYDAYFN 1812
Db 1099 -----QFPLILGSPEDMDLEELIQDLSRLQRSLRHSQ-----E 1134
QY 1813 KAPYFTNEGFPNPPCTEGTKTFS-----GNEBIYIIGADNACWIPGNEDKICNGPL 1864
Db 1135 VPRPYIA-----ARFSLPAVFHPGNQKY--GGPDNRGLFPGH----- 1171
QY 1865 KPQKQLFKFRATNNGO--FTDSYSDV-----KTLGGLSERTVEIILSVTLCL 1915
Db 1172 ----RYVLFVLAVLQKNEPTFAASPSDFQLDNPDQPIVDG--BGLIWIWGPVLA 1225
QY 1916 SIIILGTALFAPARIROKKEGTYSPQDAEIIIDTKLDQLITVADL--ELKOBELTR 1972
Db 1226 FIICIVIAILY-----KNK-----PDSKRKDSERTKCLLNADLAPHPKOPVEMR 1273
QY 1973 -----PISKSFLOHVELCTNNLKFQEFSEL---PKFLQDLSSTDDADLP 2016
Db 1274 RINFQTPGMLSHPPIDTWAHEMERLKANDSKLSQEVESIDPGQOFTWEHSNLEA--- 1330
QY 2017 WNRKARFNPKYNNNNRVKLADASVPGSDYINASYGKLCPNNEFTATQGPLPGTVG 2076
Db 1331 -NPKPNRYANVIAY--DHSRVILQLEGIMGSDYINANYVDYRRQNAVYATQGPLPETFG 1388
QY 2077 DFRWMTWETRAKTLMLTQCPEKGRIRCHOYMPEDKNKPVTVGDIIVITKLMEDVQI-DWT 2135
Db 1389 DFRWVWEQRSATVNMVMTLEESRKVKCDQYWP--NRGTETYGFIQVT-LLODTMELATFC 1445
QY 2136 IRDLKIERHG--DCMTVRQCNFTAMPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCS 2193
Db 1446 VRTFSLHKSGSSEKREVRHFQFTAMPDGHVPEYPTPLAFLRRVKTCPNPDAGPVVHCS 1505
QY 2194 AGVGRGTGFALDHLTOHNDHDFDIYGLVABLRSERCMQNLAQYIFLHCQILDLLS 2253
Db 1506 AGVGRGTGCFIVIDAMLERIRTEXTVDYVGHVTLMRSQRYNYVQTEDQYSFIHEALLA 1565
QY 2254 NKGSNQPICFVNYSALQKMDSLDAMEGDVELEWE 2287
Db 1566 CGNTEVPARSL-YTYIQKLAQVEPGEHVTGMLE 1598

RESULT 15
T30938
receptor tyrosine phosphatase - medicinal leech
C:Species: Hirudo medicinalis (medicinal leech)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30938
C:Gerahon, T.R.; Baker, M.; Nitaabach, M.; Wu, P.; Macagno, B.R.
submitted to the EMBL Data Library, December 1997
A:Description: two receptor tyrosine phosphatases of the LAR subfamily are expressed in
A:Reference number: Z20939
A:Accession: T30938
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2051 <GER>
A:Cross-references: UNIPROT:O44328, EMBL:AF017083; NID:g2695654; PID:g2695655; PIDN:AA89
C:Gene: LAR2
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;
ogv

Query Match 8.6%; Score 1033; DB 2; Length 2051;
Best Local Similarity 23.3%; Pred. No. 6e-43;
Matches 465; Conservative 312; Mismatches 703; Indels 518; Gaps 90;

QY 535 RLVPFTEHMISAFITMEGPPTVLSVRTROQVPSSIKIINYKNISSILLY----- 589
Db 2 KMLPTSATLARLAFK-FQSPGSLIMTHRSRPRPVVQKVMGNMAT-----FMCVVVG 55
QY 590 -DPPE-----YPNGKITHYIY--AMELDTNRAFOITTDNSFLITGLKYYTKYKRVAA 642
Db 56 KPAEMTWYHNRKLESEFTYQOSSRKIFTHHNSLVLRVE-----PVKK--RDEGRYEC 107
QY 643 THDG-----ESSLSREND-----IFVRTSEDPESP 669
Db 108 AHNGVGVDRVSFNLVSVTKNDPIAGPPELIVHPALGSVERYSRTVMACSKGNPEPT- 166
QY 670 QDVEVIDVTADIRLKWSPPE---KPNGII-----IAYEVLYKNI--DTLYMONTSTD 718
Db 167 -----PSMLKDPPIEFTPVRFVMPVTGALQLLNSTYSYEGKYECIAENSHGVTVNSQQAT 221
QY 719 IIL--RNLRPH-----TLNIVSVRSTYRFGH 742
Db 222 LVLKARRIPPHFSALPDNAEVHGGSLNLTCIASGSEPPFIIMKRNLDLSHRS--NNHSH 280
QY 743 GNOVSSLLSVRTSETVPDSAPENITYKNISS-----GEIELSFL--PPS-----SP 786
Db 281 GTGVRL-----EGITESA--NYTCHALSELGTASHVVQVKNVLPKPPSSSLWTEVSP 332
QY 787 NGIIKKYTIYLRKSNGN-----EERTINTSITQNIKVKKYT 824
Db 333 N-----IVHLKWSPGNSDPVDSYIIRPRPRYSHPDNFTFTDVGSDTDDHIAIRLYT 386
QY 825 QHIIEVSASTLKGEVGRGAPISILTEEDAPDSPQDFSVKQLSGVTVKL----- 873
Db 387 EYEFVRYAVNKLGRSASSTPVDVVTGELVPGSEPPNVRARPVSGTIVVVTGRTNHCATG 446
QY 874 SQQPLEPENGIIILYTVY--VMNRSSIKTINVETS--LELSLDYNYVEYSAVVTASTRFG 930
Db 447 SYK-----GYKYVITGQPVVTVPSLWTLIHVENSHTLTLNLKENHYTISVLAFSTVG 500
QY 931 DCKTGSNIISFOTPEGAPSDPKPVYVYANLSSSIILFWTPPKPENGIIQYYSVYRNTS 990
Db 501 DGPLSDNVQVTRP--GVP--DQPTNPLGESVSSTSIHIQWDRPAS--NNIVRYKLRWDSNT 557
QY 991 GTFMQNFTHLHETNDFDNMTVSTIIDKLTFISYTYFWLTASTSVGNKSSDIIIEVYTDQ 1050
Db 558 ----RSNFSVDIGRKQDQRIIHWYLLTDLAASTVYHMTLSAVAEGGEGPPTT-TIQLQTHP 613
QY 1051 DIPEGFVGNLTVESTISSTAINVSWV--PPAOPNGL-----VFYVVS 1089
Db 614 YVLKSSPTFVSAKALNSICVEVTWLFPPSSSSSSSLPPLPPSSSSSSSAIVDGFNLYFST 673
QY 1090 LILQOTPRHRPPLVYTERSIYFDNLEKYTDYILKITPSTEKG---FSDTYTAQYIKTE 1146
Db 674 KNFDKSPHRSNLATSLDLSRATLHDLKFTAYKIVAGVYNRAGEPLSDV----IVKTE 729
QY 1147 EDVPTSPITNFKNLSSTVLLSWDPP-----VRPNGALISYDLTLQGPNEYSPFI 1198
Db 730 EDIPGEKRVKY--EALNSTSVHLEKAFQDGHNGDIK--GYVVFYTELDENEEEKFEPV 786
QY 1199 TSD-----NYIILBELSPFTLYSPFAAARTKGLG-PSSILFFTYDSEVPLAPPO--NL 1249
Db 787 RSDPLSSHQTEITIGLSPSTLYQOVCAVTRKGDGDKSKKVKTKAAMP--SPKGPFTL 845
QY 1250 TLINCTSDFVWLKWSPLPGGIVKYV----SFKIEHETDTIYKKNISGFKTEAKLVGL 1305
Db 846 SLMKEENPILVLTWQDPKFSGHVGIVGLVAFGVKGESHVEBERRFGNIHRTPTF----L 901
QY 1306 EPVSTYSIRVSAFTKVGNGQNSNVVKTFTQESVDP--VVONMQCMATSHQSVLVKWDPP- 1363
Db 902 EKGATYEFKIAAKNAVGYGEYATEVI--TTPDGAPSGAPQNLASAAVLGAHTIRLHWDPV 959

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 14:42:14 ; Search time 272 Seconds
(without alignments)
4313.138 Million cell updates/sec

Title: US-10-673-885-2

Perfect score: 11985

Sequence: 1 MDPLIFLPIFLFTGTQVD.....MDSLDMEGDVELEWEETM 2291

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10536	87.9	2302	088488	088488 rattus norv
2	4611.5	38.5	1086	08BY76	08BY76 mus musculus
3	1323	11.0	2029	09VIS8	09VIS8 drosophila
4	1319	11.0	2029	1 LAR DROME	P16621 drosophila
5	1190	9.9	1597	096DM3	096DM3 drosophila
6	1188	9.9	1912	1 PTPD HUMAN	P23468 homo sapien
7	1186	9.9	1889	2 Q7QX2	Q7QX2 anopheles g
8	1186	9.9	1898	2 Q9EQ17	Q9EQ17 mus musculus
9	1181	9.9	1898	2 Q64604	Q64604 r protein-t
10	1172.5	9.8	1529	2 Q6PG86	Q6PG86 mus musculus
11	1168.5	9.7	1896	2 Q9IAJ1	Q9IAJ1 xenopus lae
12	1162.5	9.7	1887	2 Q9QW67	Q9QW67 rattus sp.
13	1160	9.7	1607	2 Q8IRSO	Q8IRSO drosophila
14	1160	9.7	1767	2 Q9W4F5	Q9W4F5 drosophila
15	1158	9.7	1894	2 Q64487	Q64487 mus musculus
16	1155.5	9.6	1897	1 PTPF HUMAN	P10586 homo sapien
17	1153	9.6	1767	2 Q24495	Q24495 drosophila
18	1147.5	9.6	1948	1 PTNS HUMAN	Q13332 homo sapien
19	1144	9.5	1898	2 Q86WS0	Q86WS0 homo sapien
20	1141.5	9.5	1556	2 Q8IR87	Q8IR87 drosophila
21	1141.5	9.5	1631	1 PTP1 DROME	P35992 drosophila
22	1141.5	9.5	1631	2 Q86NN9	Q86NN9 drosophila
23	1141.5	9.5	1631	2 Q9VYV1	Q9VYV1 drosophila
24	1132	9.4	1998	2 Q8CIW2	Q8CIW2 mus musculus
25	1124.5	9.4	1788	2 Q9IAJ0	Q9IAJ0 xenopus lae
26	1113	9.3	1956	2 Q6MZFE	Q6MZFE homo sapien
27	1109.5	9.3	1904	2 Q64699	Q64699 mus musculus
28	1107	9.2	1997	1 PTPB HUMAN	P23467 homo sapien
29	1086	9.1	2200	1 LAR CAEL	Q9BMM8 caenorhabdi
30	1079	9.0	1863	2 Q64605	Q64605 rattus norv
31	1044.5	8.7	5202	2 Q6S362	Q6S362 homo sapien

32	1033	8.6	2051	2	044328	044328 hirudo medi
33	987	8.2	1064	2	08SXB2	08SXB2 drosophila
34	902	7.5	1406	2	09W6V5	09W6V5 gallus gall
35	893	7.5	2222	2	07QEG7	07QEG7 anopheles g
36	877.5	7.3	1102	2	080VN7	080VN7 mus musculus
37	870	7.3	2169	2	08AV58	08AV58 gallus gall
38	867.5	7.2	1267	2	089945	089945 gallus gall
39	864	7.2	1254	2	08V8V0	08V8V0 mus musculus
40	863	7.2	1238	1	PTFJ MOUSE	PTFJ MOUSE mus musculus
41	861	7.2	1238	2	08K3Q2	08K3Q2 mus musculus
42	860	7.2	1238	2	08CIW9	08CIW9 mus musculus
43	859.5	7.2	1198	2	09JLU0	09JLU0 mus musculus
44	855.5	7.1	1226	2	07TSY7	07TSY7 mus musculus
45	855	7.1	1367	2	Q20120	Q20120 caenorhabdi

ALIGNMENTS

RESULT 1

088488 ID 088488 PRELIMINARY; PRT; 2302 AA.
AC 088488;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glomerular mesangial cell receptor protein-tyrosine phosphatase
DE precursor.
GN Name=PTPRQ;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wislar;
RX MEDLINE=98395110; PubMed=9727007; DOI=10.1074/jbc.273.37.23929;
RA Wright M.B., Hugo C., Seifert R., Disteche C.M., Bowen-Pope D.F.;
RT "Proliferating and migrating mesangial cells responding to injury
RT express a novel receptor protein-tyrosine phosphatase in experimental
RT mesangial proliferative glomerulonephritis.";
RL J. Biol. Chem. 273:23929-23937(1998).
DR EMBL; AF063249; AAC34801.1; -;
DR PIR; T14328; T14328.
DR HSSP; P10586; 1LAR.
DR GO; GO:0016787; P:hydrolase activity; IEA.
DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0004872; P:receptor activity; IEA.
DR GO; GO:0008470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR000387; TYR phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00041; fn3; 16.
DR Pfam; PF0102; Y_phosphatase; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00700; PRTYPHPHTASE.
DR SMART; SM00060; FN3; 16.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00853; FN3; 14.
DR PROSITE; PS00383; TYR PHOSPHATASE 1; 1.
DR PROSITE; PS00056; TYR PHOSPHATASE 2; 1.
DR PROSITE; PS00055; TYR PHOSPHATASE_PTP; 1.
KW Hydrolase; Receptor; Signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 2302 glomerular mesangial cell receptor
FT protein-tyrosine phosphatase.
SQ SEQUENCE 2302 AA; 256822 MW; F0FA703022EB25D5 CRC64;

Query Match 87.9%; Score 10536; DB 2; Length 2302;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;


```
Db 2161 VROCNETGPERGVPENTTFLIHFKLVKRTSRADHTPMWVHCSAGVGRGTGVFIALDHLT 2220
QY 2210 QHINDHFDVYIGLVAELSRMCMQNLQAQYIFLHQICLDLNSKGSNQCIFVNSAL 2269
Db 2221 QHINHDVYIGLVAELSRMCMQNLQAQYIFLHQICLDLNSKGGHQPVCVFNYSYL 2280
QY 2270 QKMSLDAMEGDGVELEWEETM 2291
Db 2281 QKMSLDAMEGDGVELEWEETM 2302

RESULT 2
Q8BY76 PRELIMINARY; PRT; 1086 AA.
AC Q8BY76;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A630028F16 product:hypothetical fibronectin type III
DE domain/fibronectin type III repeat containing protein, full insert
DE sequence. (fragment).
CN Name=Ptpqr;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Hanagaki T., Hara A., Hashizume W.,
RA Fukuda S., Furuno M., Anamitsu K., Hiraoka T., Hirozane T.,
RA Hayashida K., Hayatsu N., Hiramoto K., Kondo S., Konno H., Koyama T.,
RA Hori P., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki R., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi P., Takaku-Akai H., Tanaka Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK01657; BAC31021.1; -.
DR HSSP; P40189; 1BQU.
DR MGD; MGI:1096349; Ptpqr.
DR GO; GO:0042472; Primer ear morphogenesis; IMP.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003534; Royaljelly.
DR Pfam; PF00041; fn3; 9.
DR PRINTS; PRO0014; FNTPEIII.
DR PRINTS; PRO1366; ROYALJELLY.
DR SMART; SM00060; FN3; 9.
DR PROSITE; PS00853; FN3; 7.
DR Hypothetical protein.
FT NON TER 1086 1086
SQ SEQUENCE 1086 AA; 120855 MW; 84A18CF2F6857F3C CRC64;

Query Match 38.5%; Score 4611.5; DB 2; Length 1086;
Best Local Similarity 83.2%; Pred. No. 5.5e-233;
Matches 877; Conservative 84; Mismatches 90; Indels 3; Gaps 3;

QY 1 MDPLIFLLFLTGTSTQVDVNVVPGTRYDITISIS-TTYSPTVTRITPNTVTPGPP 59
Db 1 MDPLFFFLSLTGTSTQVDVNVVPGTRYDITISIS-TTYSPTVTRITPNTVTPGPP 60
QY 60 VFLAGERVGSAGILLSSWNTPPNPNGRILISYVVKYKEVCPWMTQVTVQVSKPDSLEVL 119
Db 61 VFLAGERVGSAGILLSSWNTPPNPNGRILISYVVKYKEVCPWMTQVTVQVSKPDSLEVL 120
QY 120 NLNPGTTYEIKVAENSAGIGVFSDFPFLQTAESAPGVVNLTVYAYNSAVKLIWYLP 179
Db 121 NLNPGTTYEIKVAENSAGIGVFSDFPFLQTAESAPGVVNLTVYAYNSAVKLIWYLP 180
QY 180 QPQKITSPKISVKGARSGIVVKDVSIRVEDILTKGLPECNENSESFLMSTASPSPTLGR 239
Db 181 QPQKITSPKISVKGARSGIVVKDVSIRVEDILTKGLPECNENSESFLMSTASPSPTLGR 240
QY 240 VTPPSRTTHSSSTLTQNEISSVWKEPISFVWTHLRPTTYLFEVSATTEAGVIDSTIVR 299
Db 241 ATPPLPETHLSNLTLARNKISSVWKEPISFVWTHLRPTTYLFEVSATTEAGVIDSTIVR 300
QY 300 TPESVPEGPQPCVNTGNITGKFSILWDPPTIVTGKFSYRVLYGSPGRILDNSTKDLKF 359
Db 301 TPESVPEGPQPCVNTGNITGKFSILWDPPTIVTGKFSYRVLYGSPGRILDNSTKDLKF 360
QY 360 AFTNLTPFTMYDVYIAAETSAGTGPKNISVFTPPDPVPGAFDQLQAEVSTQVITWK 419
Db 361 VFTLTPFTMYDVYIAAETSAGTGPKNISVFTPPDPVPGAFDQLQAEVSTQVITWK 420
QY 420 PRQNGIINOYRVKVLVPTGTLILENTLTGNNEYNIDPKAPEIIVNVEPMVGLYEGSAE 479
Db 421 PRQNGIINOYRVKVLVPTGTLILENTLTGNNEYNIDPKAPEIIVNVEPMVGLYEGSAE 480
QY 480 MSSDLHSLATFTYNSHPDKNFPARNRAEDQTSF-VVYTRNQYITDIAAQSLSVIRRLVP 538
Db 481 MSSDLHSLATFTYNSHPDKNFPARNRAEDQTSF-VVYTRNQYITDIAAQSLSVIRRLVP 539
QY 539 FTEHMISVSAFTMGSGPPTVLSVRTRQOVPSISIKIINYKNISSSSILLYWDPPYPNGK 598
```

Db 540 FTEHTSVSAFTVMSGPTTLVTRREQVPSQIINKYSSSSILLYWDPPEYNGK 599
Qy 599 ITHYTYAMELDTNFAQTITDINSPLITGLKKTYYKRWVAASTHDGSSISENDIFV 658
Db 600 ITHYTYAMELDTNFAQTITDINSPLITGLKKTYYKRWVAASTHDGSSISENDIFV 659
Qy 659 RTSDEPSSPDQVEVIDYADEIRLKWSPPKPKNGIIIAEYLYKNIIDTLTKNTSTTD 718
Db 660 RTEDEPSSPDQVKYTDVSPSELSLTWSPKPKNGIIIAEYLYKNIIDTLTKNTSTIN 719
Qy 719 IILNRPHTLNINISVRSYTRRGHGNQVSSLSVTSRTSETVPSAPENIYKNISSGEIL 778
Db 720 ITLSDLKPYTLNISTQSYTRLGHGNQSSLSVTSRTSETVPSAPENIYKNISSGEIEI 779
Qy 779 SFLPPSPNGIIKQYTYILKRSNGEERTINTSLTQNLKVLKKTYYLIEVSASTLKE 838
Db 780 FFLPPSPNGIIKQYTYILKRSNGEERTINTSLTQNLKVLKKTYYLIEVSASTLKE 839
Qy 839 GYRSAPISILTBEDAPSPQFQSVKQLSGVTWKLQSWPPLPENGIIILYTYVWNRSSL 898
Db 840 GYRSAPISILTBEDAPSPQFQSVKQLSGVTWKLQSWPPLPENGIIILYTYVWNRSSL 899
Qy 899 KTNVTETSLSDLDYVVEYSAYVTASTRFGDKTGSNIISFQTPGAPSPDPKDYVA 958
Db 900 KTNANEVSLSDLDYHADYSAYVTASTRFGDKTGSNIISFQTPGAPSPDPKDYVA 959
Qy 959 NLSSSSIIILFWPPSPKNGIIQYISVYNTSGTFMONTLHETLNDPNTVSIIDKL 1018
Db 960 NLSSSSIIILFWPPSPKNGIIQYISVYNTSGTFMONTLHETLNDPNTVSIIDKL 1019
Qy 1019 TIFSYTFLWTASTSGNGKSSDIIEVYTDQI 1052
Db 1020 AVFSYTYTFLWTASTLGVNGKSSDIIVHTYTDQI 1053
Qy 1020 AVFSYTYTFLWTASTLGVNGKSSDIIVHTYTDQI 1053
ID Q9V1S8 PRELIMINARY; PRT; 2029 AA.
AC Q9V1S8;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE CG10443-PA.
GN Name-Lar: ORFNames=CG10443;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abrial J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Yang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003663; AAF53837.3; -.
DR HSP; P10596; ILAR.
DR FlyBase; FBGN0000464; Lar.
DR GO; GO:0004723; P:protein tyrosine phosphatase activity; IDA.
DR GO; GO:0007155; P:cell adhesion; IMP.
DR GO; GO:0008045; P:motor axon guidance; IMP.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IDA.
DR GO; GO:0008360; P:regulation of cell shape; IMP.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.

Db 1451 HPIPISEFANHIERLKNQKFSQESIEPG-QQFTWDSNLSHNKSKRYANVTAY 1509
 QY 2031 NNNNRVKLTADSVPSGSDYINASYISGYLCPNEFIATQGLPGTCDVFRWVWETPAKTL 2090
 Db 1510 -DHSRVQLPAVEGVGSDYINANYCNGYRKNAYATQGLPGTCDVFRWVWETPAKTL 1568
 QY 2091 VMLTQCFERGRIRCHOYWPEDNKPVTVFQDIVITKLMEVDQIDTIRDLKIERHG--DCM 2148
 Db 1569 VMTRLEERTRIKCDQWP--TRGTETYGOIFVITITELATYSIRTFQLCRQGNDR 1626
 QY 2149 TVRQCNTAWPBGVSPENSAPLHFVKLVASRAHDTTMMVHCSAGVGRTGVFIADHL 2208
 Db 1627 EIKQLQFTAWPDGVDHPAPFLQFLRCRALTTPESGPVHCSAGVGRTGCVIIVDSM 1686
 QY 2209 TQINDHDFVDYIGLVAELRSEKCMQVLAQVIFLHCILLLLSKNGSNQIPCFVN--- 2265
 Db 1687 LERMKEHEDYGHVTVCLRAQRYNVQTEDQYIFTHDALEAI-----ICGVTRVP 1738
 QY 2266 ----YSALQKM-----DSLDAWE 2279
 Db 1739 ARNLTHLQKLITEGETISGME 1762

RESULT 4

LAR_DROME STANDARD; PRT; 2029 AA.
 AC F16621;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Protein-tyrosine phosphatase Lar precursor (EC 3.1.3.48) (Protein-tyrosine-phosphate phosphohydrolase) (DLAR).
 DE Names-Lar;
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90046860; PubMed=2554325;
 RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
 RT "A family of receptor-linked protein tyrosine phosphatases in humans and Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=96178473; PubMed=8598047; DOI=10.1016/S0092-8674(00)81036-3;
 RA Krueger N.X., van Vactor D., Wan H.T., Gelbart W.M., Goodman C.S., Saito H.;
 RT "The transmembrane tyrosine phosphatase DLAR controls motor axon guidance in Drosophila.";
 RL Cell 84:611-622(1996).
 CC -!- FUNCTION: Possible cell adhesion receptor. It possesses an intrinsic protein tyrosine phosphatase activity (Pirase). It controls motor axon guidance.
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein tyrosine + phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Selectively expressed in a subset of axons and pioneer neurons in the embryo.
 CC -!- SIMILARITY: Belongs to the protein-tyrosine phosphatase family. Receptor class 2A subfamily.
 CC -!- SIMILARITY: Contains 9 fibronectin type III domains.
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; M27700; AAC28668.1; -;
 DR EMBL; U36857; AAC47002.1; -;
 DR EMBL; U36849; AAC47002.1; JOINED.
 DR EMBL; U36850; AAC47002.1; JOINED.
 DR EMBL; U36851; AAC47002.1; JOINED.
 DR EMBL; U36852; AAC47002.1; JOINED.
 DR EMBL; U36853; AAC47002.1; JOINED.
 DR EMBL; U36854; AAC47002.1; JOINED.
 DR EMBL; U36855; AAC47002.1; JOINED.
 DR EMBL; U36856; AAC47002.1; JOINED.
 DR PIR; A36182; TDFELK.
 DR HSP; P10586; ILAR.
 DR FlyBase; Fgnm000464; Lar.
 DR GO; GO:0004725; P:protein-tyrosine-phosphatase activity; IDA.
 DR GO; GO:0008045; P:motor axon guidance; IMP.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IDA.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003962; FNIII-subd.
 DR InterPro; IPR007110; IG_c2.
 DR InterPro; IPR003598; IG_c2.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00441; fn3; 9.
 DR Pfam; PF00447; ig; 3.
 DR Pfam; PF0102; Y_phosphatase; 2.
 DR PRINTS; PR00104; FNTYPEIII.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR SMART; SM00060; FN3; 9.
 DR SMART; SM00408; IGC2; 3.
 DR SMART; SM00194; PTPC; 2.
 DR PROSITE; PS00853; FN3; 9.
 DR PROSITE; PS00835; IG_LIKE; 3.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
 KW Cell adhesion; Glycoprotein; Hydrolase; Immunoglobulin domain; Protein phosphatase; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 32
 FT CHAIN 33 2029 Protein-tyrosine phosphatase Lar.
 FT DOMAIN 33 1377 Extracellular (Potential).
 FT TRANSMEM 1378 1402 Potential.
 FT DOMAIN 1403 2029 Cytoplasmic (Potential).
 FT DOMAIN 36 128 IG-like C2-type 1.
 FT DOMAIN 140 224 IG-like C2-type 2.
 FT DOMAIN 324 316 IG-like C2-type 3.
 FT DOMAIN 322 410 Fibronectin type-III 1.
 FT DOMAIN 416 508 Fibronectin type-III 2.
 FT DOMAIN 515 603 Fibronectin type-III 3.
 FT DOMAIN 610 704 Fibronectin type-III 4.
 FT DOMAIN 709 807 Fibronectin type-III 5.
 FT DOMAIN 812 904 Fibronectin type-III 6.
 FT DOMAIN 909 1002 Fibronectin type-III 7.
 FT DOMAIN 1006 1099 Fibronectin type-III 8.
 FT DOMAIN 1101 1203 Fibronectin type-III 9.
 FT DOMAIN 1492 1738 Protein-tyrosine phosphatase 1.
 FT DOMAIN 1781 2029 Protein-tyrosine phosphatase 2.
 FT ACT_SITE 1670 1670 Phosphocysteine intermediate (By similarity).
 FT ACT_SITE 1961 1961 Phosphocysteine intermediate (By similarity).
 FT DISULFID 57 111 Potential.
 FT DISULFID 161 209 Potential.
 FT DISULFID 256 301 Potential.
 FT CARBOHYD 176 176 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 253 253 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 298 298 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 553 553 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 616 616 N-linked (GlcNAc...) (Potential).

FT	CARBOHYD	666	666	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	721	721	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	774	774	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	915	915	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	962	962	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	1183	1183	N-linked (GlcNAc. . .)	(Potential).
FT	CARBOHYD	1304	1304	N-linked (GlcNAc. . .)	(Potential).
SO	SEQUENCE	2029 AA;	229027 MM;	536A0C794D3DC800	CRC64;

Query Match 11.0%; Score 1319; DB 1; Length 2029;
 Best Local Similarity 24.3%; Pred. No. 6.3e-60;
 Matches 530; Conservative 292; Mismatches 762; Indels 600; Gaps 85;

Qy	270	VTHLRPYTYLFEVSAATTEGYDSTIV	---RTPESVPEGPPQNCVTGNITGKSPFILM	326
Db	5	MTAARPIAALSLLVLSLLT	---WTHPTVDAAHPEIIRKPFQNGVRGVGVASPYCAARG	61
Qy	327	DPPTI	-----VTGKFSYRVELYGPSG	358
Db	62	DPPPSIVVRKMGKVGSTQSRVTVLEQPGGSIILRIEPVRAGRDDAPYECVAENGVDVAV	121	
Qy	359	FAFTNLTPFTMYDYVYIAAETSAGTGPKSNISVFTPPDPVPGAVFDQLAAVESTQVRITWK	418	
Db	122	SADATLT	-----IYEGDKTPAG	162
Qy	419	KPROPGIINQYRKVLVPETGIIENLTLLTGNEYINDMPAIEVINIVPWCGLYEGSA	478	
Db	163	AIGNPTP	---NIYWIK---NQTKVDMSPRYSLKDGLF	211
Qy	479	EMSSDL	---HSLATFIY---NSHPDKNFPARNRAEDQTSVVTTRNOYITDIAAELQSVVIR	534
Db	212	ENSMGTEHSKATNLYVKVRVP	-----PTFSPPTTISEVMLGSLNLSCTA	260
Qy	535	RLVPFTEHMSVSAPTINGEGPPTVLSVTRQQVPSSIKI	INYNKNISSSSILLYWDPPPY	594
Db	261	SPMHPVXWVKGSEDLT	-----PENEMPIGRNVQLQLIQESA	297
Qy	595	PNGKIHTHYIYAMELDTNRAFOITIDNSPLITGLKKYTKYKRVAASTHGDSSISEEN	654	
Db	298	-----NYTC	-----IAAST	311
Qy	655	DIFVRTSEDEPSSPDQVEVIDVTADBIRLKWS	---PPEKPGIIIIAYEVLYKNIDTLMYK	712
Db	312	SVSVVKQSLP	---TAPTDVQISEVTATSVRLWSYKGPEDLQYVYQKPNANQAFSEIS	370
Qy	713	NTSTDDIILNRLPHLLYNI	SVRSYTRFGHGNQVSSLLSVRTSTVPDPSAPENITYKNIS	772
Db	371	GLIITMYVVRALSPTYTEFVYIVAVNNIGRC	---PPSAPATCTTGETKMSAPRNVQVTRLS	429
Qy	773	SGETLSFLPPSSPNGIIKKYTIYLRKSGNEERTINT	---TSLTONIKVLKKYCYII	828
Db	430	SSTWITWEPPETPNGQVTGKYVYTNSNOPEASWNSQWVDSBELTIVSDVTPHALYTV	489	
Qy	829	EVSASTLKGCVRSAPISIIITEEDAPDSPPODFSVKOLSGVTVVKLSWQPLEPNGIILYY	888	
Db	490	RVQAYTSMGAGPMSTPVQVKAQQGV	---SQPSNFRATDIGETAIVTLQWTKPTHSENIVHY	548
Qy	889	TVYVW	-----NRSSLKTIINTVETLSLELSDLYNVEYSAYVASTRFGDKTGSNIISFQT	943
Db	549	ELY	---WNDTYANQAAHHKRAINS	605
Qy	944	PEGAPSPDKDVIYANLSSSSIIILFTWPP	---SKPNGIIIOYSVVY	996
Db	606	KQYVPGAGPRNITAIATSSITISLSWLP	PPVVERNGRIIYKYKFFVEVGHEDDEATM	663
Qy	997	FTLHELNDPDMTVSTIIDKLIATIPSYTYTFLWTASTSVGNKNSSDII	EVYTDDQIDPEGF	1056
Db	664	---TL	-----NMT	711
Qy	1057	VGNLTYESISSTAINVSWVPPAQP	---NGLVPFYVYSLIIQOTPPHVRPPLVTYE	1108
Db	712	PQDYKATPLNASTSHVSKWPKLEKDRNGIRNGY	-----HIHAQELRDEKGFLE	761

Db 1569 VNMTRLEERTRIKCDQWP--TRGTETYQIFVTITQELATYSIRTFQLCRQGFNDRR 1626
Qy 2149 TVRQCNFTAPBHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVCRGTGFIALDHL 2208
Db 1627 EIKQLQFTAMPDGHGPDHPAPFLQFLRRCAUPTPESGFIIVHCSAGVCRGTGCIVIDSM 1686
Qy 2209 TQIHNDHDVDYIYGLVAELSRMCMVQNLQAVIFLHOCILDLNLSKGSNQPICFVN--- 2265
Db 1687 LERMKHEKIIDYGHVTCURAGNYVQTEDQYIFIHDAILEAI-----ICGVTEVP 1738
Qy 2266 ----YSALQKM-----DSLAME 2279
Db 1739 ARNLHQLKLLITERGETISGME 1762

RESULT 5

Q960M3
ID Q960M3 PRELIMINARY; PRT; 1597 AA.
AC Q960M3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE LD45391p.
GN Name=Lar;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051985; AAK93409.1; -
DR HSSP; P10586; 1LAR.
DR FlyBase; FBgn000464; Lar.
DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IDA.
DR GO; GO:0007155; P:cell adhesion; IMP.
DR GO; GO:0008045; P:motor axon guidance; IMP.
DR GO; GO:0008470; P:protein amino acid dephosphorylation; IDA.
DR GO; GO:0008360; P:regulation of cell shape; IMP.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_PP.
DR Pfam; PF00041; fn3; 8.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00014; ENTPEIIII.
DR PRINTS; PR00700; PRTYPHPHTASE.
DR SMART; SM00060; FN3; 8.
DR SMART; SM00194; PTEG; 2.
DR PROSITE; PS00853; FN3; 8.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_FTP; 2.
KW Hydrolase.
SQ SEQUENCE 1597 AA; 182017 MW; A8C2B41445636800 CRC64;

Query Match 9.9%; Score 1190; DB 2; Length 1597;
Best Local Similarity 24.4%; Pred. No. 2 5e-53;
Matches 420; Conservative 228; Mismatches 555; Indels 518; Gaps 61;
Qy 684 LKWSPEKPGNIIAYEVLY-----KNIDTLAMKNTSTTDIILRNLRPHLNLNISVR 735
Db 3 ITWEPETENGQVTKYVYTTNSNQPEASWNSQVMDNSLITV--SELTPLAIYTVRQ 60

Qy 736 SYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIELSFLPSSPNIKKYTI 795
Db 61 AYTSNGAG-PWSTPVQVKAQGV--SQSPNPRATDIGETAVTLQWTKPSSHSENIHVYEL 118
Qy 796 YLKRNGNE---ERINTTSLTONIKVLKYYTOYIIEVSASTLKEGVRSAPISILTEED 852
Db 119 YWMDTYANQAHHKRISNEAYT--LDGLYPDLYIWLAAARSQREGGATTPPIPVRTQY 176
Qy 853 APDSPQDPFVKQLSGVTVKLSWQPP--LEPNGIILYTVVYV-----NRSSLKTINVT 904
Db 177 VEGAPRRIATAISTTISLSWLPFPPVERSNGRIIYKVFVEVGREDDAATWTLMN- 235
Qy 905 ETSLESLDLDNVNVEYSAYVTASTRFGDGKTSNIIISFOTPEGAPDPKDVVYANLSSS 964
Db 236 -TSIVLDELKRWTEYKIWLACTSVGDGPRSHPII-LRTQEDVPGD-PQDVKATPLNSTS 292
Qy 965 IILFWTPPSKP--NGIIQYISVYVYNTSGTFMQNFTLHNLNDFNMVSTIIDLKLTFS 1022
Db 293 IHVSWKPPLEKDRNGIIRGYHIH-----AQEL----- 319
Qy 1023 YVTFWLTASTSVGNGKSSDIIENVYDODIPEGFVGN-LTYESISSTAINSVWVPAQPN 1081
Db 320 -----RDEGKGFLENEPFKFDVDTLEPNVTGLQPD--- 349
Qy 1082 GLVFFYVSLILQOTPRHVRPPLVTVYERSIYFDNLEKYTDYILKITPSTEGFSDTYTAOL 1141
Db 350 -----TKYSIQVAALTRKGDGR-SAAI 371
Qy 1142 YIKTEEDVP-----ETSPINFTKNSSTSVLLSWDPPVPKNGAIIISYDL----- 1186
Db 372 VVKTPGVVPVPTVSLKIMEREPIV-----SIELEWERPAQTYGELRGYRLRWGVK 422
Qy 1187 -----TLQGNENYSITSDNYIILBELSPFTLYSPFAAARTKGLGPSILFFYDTE 1239
Db 423 DOALKEMLSGEQ-----MTKKRP---DNLEKGVYEFRAVGNHIGIGQETVKVLPQTE 474
Qy 1240 SVPLAPPQNLTLINCTSDFWMLKWSPLPGGIVKVYSFKIHEHETDTIYYKNISGFKTE 1299
Db 475 GTFPGFPSPNIT----- 485
Qy 1300 AKLVGLEPVTYSIRVSATFKVGNQGNQSNVVVKTQESVDPVQVQNMCMATSWOSVLVK 1359
Db 486 -----IRFQT-----PDV-----LCVT 497
Qy 1360 WDPKKK--ANGIIITQYMTV-----BENSTKVSPQDHWYTFIKLLANTSVEVKR 1407
Db 498 WDPPTREHRNGIITRYDVQFHKKDHGLGSENRMTL-----RKAVFTNLEENTEIFRVR 552
Qy 1408 ASTSAGEGDESTCHVSTLTPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKI-T 1466
Db 553 AYTKQAGAPPSDKLIVETERDMGRAPMSLOAEATSEQTAEIWW-EPVTSRGKLLGKIFY 611
Qy 1467 TOLRAQCKKEWSECEVYQKIYQYLAHLTEET-VYGLKKPRWYRFOVAAGTNAGYNA 1525
Db 612 TWTAVEDDLDDWQTK-----TVGLTESADLVNLEKFAQAVAIARFKNGLGLRL 659
Qy 1526 SNWISTKTLPGPPDGPENNVHVATSPFSISISWSEPAVITGPTCYLLDKVS---VDND 1581
Db 660 SEKVTVRI---KPEDVPLNLRADHVSTHSMTSLWSPPILRLT-FVNYKISFDMKVPVDSQ 715
Qy 1582 EFNISFIKSNEE-----NKTIEIKDLFIETYSVVITAFITGNISAAVVEGSSAEMIVT 1635
Db 716 GFSQTIQVPKREIILKHVYKHTHINELSPFTTYNVNVSAPSDYSY-----RPPYKIYTV 770
Qy 1636 TLESAPKDPNNMTFQKI--PDEVTKFQLTFILP-PSQPNGNIQVYQALVYREDDTFAVOI 1692
Db 771 TQMAAPQ-PMVKPDPFGVVGNGEEI---LVILPQASEEYGPISHYLVVVPED----- 818
Qy 1693 HNLIIQKNTNTVIAMLEGLKGGHTYINSVAVNSAGAPKVPMTIMDIKAPAREK-TK 1751
Db 819 -----KSNLH-----KIPDQFLTDLLPGRNKPER 843
Qy 1752 PTPYDATGKLLVSTTTITIRMPICVYSDDHGPIKNVQVLATETGAQHDGNTWKYDAYF 1811

Db 844 PNPYIA-AKFPORSIPFTF-----HLGS-----GDVH 871
Qy 1812 NKARPYTFNPFNPPCTCKTEKSGNEEYIIIGADNACMIPGNEKICNGPLKPKQYL 1871
Db 872 N-----FTNR-----KLEREKRYR 885
Qy 1872 PKFRATNMQ-----FTDSYS-----DPVKTLEGISERT 1903
Db 886 IFVRVVDTTPQKHLYTSFSPSEFLSLDMREAPPGRPHRDPNPWPAEPEVSVNRNDE-- 943
Qy 1904 VEILSVTLCLISILLGTAIFAFARIRKQEGGTVSPDARIIDTKLKLQDITVADL 1963
Db 944 PELWVVLPLWVTFVISTALIVLCVVRKRRQCKT--PDQAAVTRPLMAADLGAGPTPS 1001
Qy 1964 ELKDERLTR-----PISKKSFLQHVVEELCTNNLKFQEEFSELPKFLQDLSSTDA 2013
Db 1002 DPVDMRLNFQTMISHPPPIPISEFANHLERLKSNDNQKFSQYESIEPG-QQFTWDNS 1060
Qy 2014 DLPWNAKRNPNKIPYNNNRVKLADASVPGSDVINASYISGYICPNFEIATQGLPG 2073
Db 1061 NLEHNSKRYANVTAY-DHSRVQLPAVEGVGSDVINANYCDGYRKNAYVATQGPLOE 1119
Qy 2074 TVGDFRWVWETRAKTLVMTQCFEKRIRCHOYWPEDNKPVTVFGDIVITKLMDVQID 2133
Db 1120 TVDFWRMCMELKTATIVMTRLEERTIKCDQWP--TRGTETYGOIFVTITETQELAT 1177
Qy 2134 WTIRDLKIEHGG--DCWTVRQCNTAPWPHGVGPENSAPLIHFVKLVASRAHDTTPIVH 2191
Db 1178 YSIRTFQLCQGFNDREIKQLFTAPWDHGVDPHAPFLQFLRCALTPPSSGVIVH 1237
Qy 2192 CSAGVGRGVFIADHLHTQHINDHDVDYIYGLVAELSRMCMQVLAQYIFILHCILDL 2251
Db 1238 CSAGVGRGVYVIDSMLERMKHEKIIDYGHVTCRAQRNVYVQTDQYIFTHDAILEA 1297
Qy 2252 LSNKGSNQPICFVN-----YSALQNM-----DSLDAME 2279
Db 1298 I-----ICGVTEVPARNLTHLOKLLITEPGETISGME 1330

RESULT 6

PTPD_HUMAN STANDARD; PRT; 1912 AA.
AC P23468;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein-tyrosine phosphatase delta precursor (EC 3.1.3.48) (R-PTP-
delta).
GN Names:PTPRD;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN J. Biol. Chem. 270:6722-6728 (1995).
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
RX MEDLINE=95204468; PubMed=7896816; DOI=10.1074/jbc.270.12.6722;
RA Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT "Molecular characterization of the human transmembrane protein-
tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT alternative human transmembrane protein-tyrosine phosphatase delta
RT isoforms.";
RL J. Biol. Chem. 270:6722-6728 (1995).
RN [2]
RP SEQUENCE OF 390-1912 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases.";
RL ENBO J. 9:3241-3252 (1990).
RN [3]
RP INTERACTIONS WITH PPPIA1; PPPIA2 AND PPPIA3.

RX MEDLINE=98288299; PubMed=9624153; DOI=10.1074/jbc.273.25.15611;
RA Serra-Pages C., Medley Q.G., Tang M., Hart A., Streuli M.;
RT "Liprin, a family of LAR transmembrane protein-tyrosine phosphatase-
interacting proteins.";
RL J. Biol. Chem. 273:15611-15620 (1998).
RN [4]
RP SPLICE ISOFORM(S) THAT ARE POTENTIAL NMD TARGET(S).
RX PubMed=1479259; DOI=10.1186/gb-2004-5-2-r8;
RA Hillman R.T., Green R.E., Brenner S.E.;
RT "An unappreciated role for RNA surveillance.";
RL Genome Biol. 5:RESEARCH008.1-RESEARCH008.16(2004).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBUNIT: Interacts with PPPIA1, PPPIA2 and PPPIA3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P23468-1; Sequences=Displayed;
CC Notes=May be produced at very low levels due to a premature stop
CC codon in the mRNA, leading to nonsense-mediated mRNA decay;
CC Name=2; Synonyms=Kidney;
CC IsoId=P23468-2; Sequence=VSP_005147, VSP_005148, VSP_005149;
CC Name=3; Synonyms=Fetal brain;
CC IsoId=P23468-3; Sequence=VSP_005150;
CC -1- PTM: A cleavage occurs that separates the extracellular domain
CC from the transmembrane segment.
CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Receptor class 2A subfamily.
CC -1- SIMILARITY: Contains 8 fibronectin type III domains.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L38929; AAC41749.1; -;
CC EMBL; X54133; CAA38068.1; -;
CC PIR; A56178; A56178.
CC HSP; P10586; ILAR.
CC Genew; HGNC:9668; PTPRD.
CC MIM; 601598; -;
CC GO; GO:0005887; C:integral to plasma membrane; TAS.
CC GO; GO:0005001; P:transmembrane receptor protein tyrosine pho. . . ; TAS.
CC GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.
CC GO; GO:0007185; P:transmembrane receptor protein tyrosine pho. . . ; TAS.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003962; FN_III subd.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003598; IG_c2.
CC InterPro; IPR000387; TYR_phosphatase.
CC InterPro; IPR000242; TY_PP.
CC Pfam; PF00041; fn3; 8.
CC Pfam; PF00047; ig; 3.
CC Pfam; PF0102; Y_phosphatase; 2.
CC PRINTS; PR0014; FNTYPEIII.
CC PRINTS; PR00700; PRTYPHPTASE.
CC SMART; SM00060; FN3; 8.
CC SMART; SM00408; IGc2; 3.
CC SMART; SM00194; PTEC; 2.
CC PROSITE; PS50853; FN3; 8.
CC PROSITE; PS50835; IG_LIKE; 3.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
CC PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
CC PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Alternative splicing; Glycoprotein; Hydrolase; Immunoglobulin domain;

[illegible]

QY 1739 TWDIKAPARKTPTDIYATGKLVSTTITIRMPICYSDHGHKQVQVLATETGAQ 1798
 Db 1190 -----PYIAAHFDVLPTEFTL-----GDD-----K 1209
 QY 1799 HDGNVTWKYDAYENKARPVFTNEGPPNPPCTEGTKFSGNEEYIIGADNACMIPGNEDEK 1858
 Db 1210 HVG-----GFTN-----KQLOSGQEVFFVLA-----VMEHAEK 1239
 QY 1859 ICNGLPKKQYLKFKFRATNIMQFTDSDYDPVKTLG-----EGLSERTVEIIL 1908
 Db 1240 M-----YATSPYSDPVVSMOLDPOPIITDEEGLI-----WVV 1271
 QY 1909 SVTLCLISLILCTAIFAFARINQKQEGTYSQDAEIID-----TKLKLDLQITVADLE 1964
 Db 1272 GPVLAVVFIIICVIAILLYKRAESDSRKSSTPNNEKIPSHHPTDPVLRRL-----NFQ 1327
 QY 1965 LKDERLTRPISKKSFLQHVVELCTNNLKFQEFSELKFLQDLSTLSDADLPWNAKRF 2024
 Db 1328 TPGMASHHPPILELADHIERKANONLKFQSBYESIDPG-QQFTWEHENLENVKPKRY 1386
 QY 2025 PNKYPNNNRVKLIADASVPGSDYINASYIGYLCNPFIAITQGLPGTGVDFMRWWE 2084
 Db 1387 ANVIAY-DHSRVLLSAIEGIPGSDYVANYIDYRKQNAVIAITQGLPETFGDFRMIWE 1445
 QY 2085 TRAKTLVMTQCEKGRIRCHQWPNEDKPVTVFGDIVITKLMEDVQI-DWTIRDLKIER 2143
 Db 1446 QRSATVMMTKLEERSRVKCDQWP--SRGTETHGLVQVT-LLDVTELATYCVRTPALYK 1502
 QY 2144 HG--DCWTVROCNFTAPWPHGVPENSAPLIHFVKLVRAASHDTPPMIVHCSAGVGRTCV 2201
 Db 1503 NGSSKREVRQFTAPWPHGVPEHTPFLAFURVKTCNPPDAGPMVHVCAGVGRTCG 1562
 QY 2202 FIALDLHTQINDHDPVDIYGLVAELSRMCMQVLAQYIFLHQILDLNKSQSNQPI 2261
 Db 1563 FIVIDAMLERIKHEKTDIYGHVTLKRAQRYNMVQTEDQYFIHDALLEAVTCGTEVPA 1622
 QY 2262 CFVNTSALQKMSLDAMEGDVELEWE 2287
 Db 1623 RNL-YAVIQKLTQIETGENVTGMELE 1647

RESULT 7

Q700X2
 ID Q700X2 PRELIMINARY; PRT; 1889 AA.
 AC Q700X2;
 DT 01-WAR-2004 (TrEMBLrel. 26, Created)
 DT 01-WAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE EbiPe103 (Fragment).
 GN Name=ebiG6103; ORFNames=ENSANGG000000004649;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -! CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008980; EAA14539.1; --
 DR HSP; P18052; LP15.
 DR GO; GO:0016787; F-hydrolase activity; IEA.
 DR GO; GO:0004725; F-protein tyrosine phosphatase activity; IEA.
 DR GO; GO:0006470; P-protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR003962; FniII subd.
 DR InterPro; IPR003961; FN-III-like.
 DR InterPro; IPR008957; FN-III-like.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR000387; TYR_phosphatase.

DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00041; fn3; 8.
 DR Pfam; PF0102; Y_phosphatase; 2.
 DR PRINTS; PRO0014; FNTYPEI1.
 DR PRINTS; PRO0700; PRTYPHPHTASE.
 DR PROSITE; PS00853; FN3; 8.
 DR PROSITE; PS00835; IG LIKE; 3.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE; PS00506; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolase.
 FT NON_TER 1
 FT NON_TER 1889
 SQ SEQUENCE 1889 AA; 213021 MW; 531EFEBD4C13ED86 CRC64;
 Query Match 9.9%; Score 1186; DB 2; Length 1889;
 Best Local Similarity 23.1%; Pred. No. 5.3e-53;
 Matches 488; Conservative 282; Mismatches 726; Indels 618; Gaps 78;
 QY 301 PESVPEGPPQNCVTGNITGKSFSLWDP-PTIVTGKFSYRVLYGSGRILD-NSTKDLK 358
 Db 2 PEIIKKPVNQVRVGAVATFFCAARGDQPTVIMRKNKSKIMMOTOSRYTVYDANGVSLR 61
 QY 359 P----AFTNLTPFTMYDVVIAAETSAG--TGPKSNISVP---TPDPV----- 397
 Db 62 IEPVRAGRDDAPYE-----CVAENGVDGAVSAEALITYEGDKTPAGFPVVKQPPTRVI 116
 QY 398 --GAVFDLQLAEVESTQVRITWKPKQPNNGIIINQYRVKLVLPETGIIILENTLLTGNNYI 455
 Db 117 ETGHTAVMOCKATGSPPKIYWLK-----DMKRVDMTNPRYSINSGSLQIDNSE-- 166
 QY 456 NDMMAPEIIVIVPEVMVGLYEGSAEWS-SDLHSLATFIY-----NSHPDK----- 498
 Db 167 -----ESDMGKYCEVAENSIGTEHTKPTLYVKVRVPPTFSRPPPEVYEVML 214
 QY 499 --NPPARNAEDQTSPPVTRNOYITDIAAEQLSVVIRLVPPTTEHMISSAFTINGEGP 556
 Db 215 GANLTLCVAGSPMPSVKRKGVDQDLPENDVPVGRNVLELDIRVSTN-YTCAQSS 273
 QY 557 PTVL---SVTRQQVPSSIKIINYKNISSSILLYWDPPPEYNGKITHYTIYAMELDTNR 613
 Db 274 LGVIEATSLVKVQSLPAAPTDVISEVTATQVRLW---SYKGPEDLQY--YVQYKPKN 328
 QY 614 AFQITTDINSFLIT-----GLKKYTKYKMRVAASHDGESSLSENDIFVRTSEDEPS 667
 Db 329 ANQASIVEISGIITMFYVVRTLSPTYEYEFYVIAVNNIGRGPPS--LPATTTTGETEMES 386
 QY 668 SPQDVEVDVTADEIRLAKSPBEKNGIIIAEVLY-----KNIDTLYMKNTSTDI 719
 Db 387 APRNIEVKPLSSSTWLTITWEPPTPNQVGTGYKYVYTTNPNQPEASWDSQMETNDMTT-- 444
 QY 720 ILRNLRPHTLNIVSVTRFGHGNQVSSLSVRSRTSETPDSAPENITKNISSGRIELS 779
 Db 445 -ISELTPHAIYIRVOAFTSKAG-PLSNPVQVKAQGV-PQPSNFRATDVEITAVLIQ 501
 QY 780 FLPPSSPNGIICKYTYILKRGNGNE---ERTINTSLTONIKVLKKYTOYIIIEVSASTLK 836
 Db 502 WSRPTHSGENIVHYELYNWDTVANEQHQRIPTETYT--LSSLYPDTLYYFWLSARSQR 559
 QY 837 GEGVSAISILTEEDAPDSPPQDSVKQLSGVTYKLSWQPP--LEPNGIILYYTYV--- 891
 Db 560 GEGATTPPIPVRTKQYVFCAPPNRYVTEATSPSTINVSMLPVPVRSNGAIYVYKVPFVE 619
 QY 892 VMNRSSLKTI-NVTETSLSDLDYVVEYSAVVYASTAFRGDCGKTSNIIISFOTPEGAPSD 950
 Db 620 VGRDSEATVTTLNSTISVLDLKKWTEYKIVLWLAGTSVGDGPR-SYPYVTRHEDVDPD 678
 QY 951 PKQVYVYANLSSSSIIILFWTPPSKE--NGIIQYYSVYVYRNTSGTFMQNFTLHETNDFN 1008
 Db 679 -PDVKASPVNSTTIIVTWKPKKRNNGIIRGYHIH-----VQEMKEE--- 721
 QY 1009 MTVSTIIDLKTIPTSYTTFWLTASTSVGNKNSDIIIEVYTDQDIEGFGVGNUTYRESIST 1068

Db 722 -----GKG-----LLNEPMKFDVVDGLEYNVT-----743
Qy 1069 AINVSVPAPQNGLVFYVYVLSLIQTPRHRVPLVTVYERSIVFDNLEKVTYDILKITS 1128
Db 744 -----GLQDPDKYSVQVVAL 758
Qy 1129 TEKGFSDDTYTAQIYIKTEEDVETSPIINTFNLS-----STSVLLSWDPVPKNGAI-----1181
Db 759 TRKGDGR-SAPISVKTGGVP-IRETV-TLKILERDPTVSI ELEWERPRQTYGELGRYR 815
Qy 1182 ISYDLTLQPNENYSFITSNDYIILELSPFTLYSPFAARTRKGLGPSIILFFTYDES 1241
Db 816 VRWGVEQALNEILQGTQLNKRINDLGRGEYEFVRVAGMNHIGIGQEAHVHLQTPGS 875
Qy 1242 PLAPPQNLTLINCTSPFVWLKWSPLPGGIKVKYSFKIHEHET-----DIIYKNIS- 1294
Db 876 PTGPPTGIAVRFPDVCITWEP-----PTREHNGOITRYDVOFHKKIDH 922
Qy 1295 GFKTE-----AKLVGLEPVSYSIRVSAFTKVGNGNOFSNVVKFTQESVDPVQNMQ 1347
Db 923 GLGTERNTVRKAVFNLDSEYIIVRAYTKQAG-PFSEKVIATERDMGRAPFSVQ 981
Qy 1348 CMATSQSVLVKWDPPKANGIITQVMVTVERNSTKVSQDHYTFIKLANTSYVFKVR 1407
Db 982 AVATSEQTVEVWEP-----996
Qy 1408 ASTSAGEBDECHVSTLPTVPSVPTNIAPSDVQSTSATLTWIRPDTILGYFQNYKI-T 1466
Db 997 -----VPS-----RGKLVGYKIFY 1010
Qy 1467 TQLRAQCKEWESECEVYQIQLYEAHLTET-VYGLKFRWRVQVAASTNAGYNA 1525
Db 1011 TMTAVEDLEWQTK-----VVGWTESADLINEKPAQYAVAAVAKYTKGLK 1058
Qy 1526 SNWISTKTLPDGPDPENHVHVSATPFSISWSBPVITGPTCYLIDVKSVDNDEFNI 1585
Db 1059 SE-----KATVKAAPQPMVQDFYGV-----VNGEIOV 1088
Qy 1586 SFIKSNEENKTIKDLIEFTRYSVITFTGNISAAVYEGKSSAEMIVTLESAPKDP 1645
Db 1089 ILPQASEE-----YGPISHY-----LIV-----VEDKA 1113
Qy 1646 NNMTFOKIPEVTKFQLTFLPPSOP---NGNIQVQYALVREDPTAVOI-----HNLS 1696
Db 1114 N---LHKIDQ---FLTEBLLSPRPKLDRLNAPYIAAMFLORNIPTFHLGSETWNP- 1166
Qy 1697 IIQKNTFTVIAMLEGLGHTYINISYAVNSAGAGPKVPMRITMDIKAPARKTKPTTY 1756
Db 1167 ---TN-----KLERGKRYIFVRV-----V 1185
Qy 1757 DATGKLLVSTTITIRMPICYSDRGPKNVQVLTATETGAQHDGNTWYDAYFNKARP 1816
Db 1186 DTPQKHLTSSPPS-----EFLALDMKEATGE-----PP 1215
Qy 1817 YFTNEGFPPPP-----CTEGKTFSGNEE---LYIIGADNACMIPGNEKICNGPLKPK 1868
Db 1216 HRPN---PNVPQTEIIIRGK-----NEEPMGLWVG-----1244
Qy 1869 QYLKFKFRATNIMGQFTSDSYSDPKVTLGELGSERTVEIILSVTLCLISILLGTATFAFA 1928
Db 1245 -----PI-----IAALVLSVCLF---VVFVR 1263
Qy 1929 RIRQKQEGGTYSPQAEIIDTKLDQLITVADLELKERLTR-----PISKKS 1978
Db 1264 RRRQPCK-----APDQAAVTRPLMAADLGAGPAPTDPVDMRRINFQTPGMISHPPISIA 1318
Qy 1979 FLOHVELCTNNLKFQEESEL---PKFLQDLSSTDAADLPWRNKRNPENKPYNNNR 2035
Db 1319 LANHVERLKANDLKFQESTESTEPGQOFTWDSHNEV-----NKPKNRYANVTISY-DHR 1373
Qy 2036 VKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRMVWETRAKTLVMTQ 2095
Db 1374 VILPPIEGVGSYINANYCDGYRKNAYVATQGLQETFGDFWRMCWELKSSTIVMTIR 1433

Qy 2096 CFEKGRIRCHOYWPBDNKFTVTFGDIVITKLMEDVQIDWTIRDLKIERHG--DCMTVRQC 2153
Db 1434 LEERSRIKCDMYWPA--RGTEVYGAMTVTITETQELATVSIIRTFQIYRNGSNERREIKQL 1491
Qy 2154 NETAPEPGEVSPNSAPLIHFVKLVASRAHDTTPMIVHCSAGVGTGVFIADLHLOHIN 2213
Db 1492 QTAWEDEGVDPHPAPFLQFLRRTKSLTPSESGPIIVHCSAGVGTGCVIVIDSLMRMK 1551
Qy 2214 DHDFVDIYGLVAELSERMCMVQNLAOXYIFLHOCILDLNLSNGSNQPICFVNVYSALOKMD 2273
Db 1552 YEKTIIDYGHVTVCLRAQRNVMQVTDQYFIHDALLEAVI-CGSTEVPARSLHNIQKLM 1610
Qy 2274 SLDAMEGDEVELEWE 2287
Db 1611 QTEPHENITGMEWE 1624

RESULT 8

Q9EQ17 PRELIMINARY; PRT; 1898 AA.
AC Q9EQ17;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Tyrosine phosphatase LAR.
GN Name=Ptpri;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=21135493; PubMed=11241288;
DOI=10.1003/1521-4141(200103)31:3<832:;AID-IMMUB32>3.0.CO;2-D;
RA Tarszowski G., Jankowski A., Hendricks W.A.J., Rolink A.G.,
Kisielow P.;
RT "Within the hemopoietic system, LAR phosphatase is a T cell lineage-specific adhesion receptor-like protein whose phosphatase activity appears dispensable for T cell development, repertoire selection and function."
RL Eur. J. Immunol. 31:832-840 (2001).
DR EMBL; AF300943; AAG40194.1; -.
DR HSSP; P10586; LLAR.
DR MGD; MGI:102695; Ptpri.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR000387; TYR phosphatase.
DR InterPro; IPR000242; Tyr_PP.
DR Pfam; PF00041; fn3; 7.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PRO0014; FNTPFIII.
DR PRINTS; PRO0700; PRTYPHPTASE.
DR SMART; SM00060; FN3; 8.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS50853; FN3; 8.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00566; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
SQ SEQUENCE 1898 AA; 211504 MW; BFD48DD11B352A4A CRC64;

Query Match 9.9%; Score 1186; DB 2; Length 1898;
Best Local Similarity 24.4%; Pred. No. 5.3e-53;
Matches 477; Conservative 282; Mismatches 713; Indels 484; Gaps 71;

QY 461 PEIVNVEPMVGLYEGSAEMSSDLHSLATFIYNSHPDKNPPARRAEDQTSPPVT----- 515
DB 33 PVFVKVPEDQTLGSEG-----VAFVC-----QATGEPKPRITWMKG 70
QY 516 --TRNOYITDIAAEOQLSVIRLVPF-----TEHMSVSAFTIMGE-PPPTVLSVRTRQOV 568
DB 71 KKVSSORFVEIFDDGAGSVLRLOPLRVORDEALIECTATNSLGEINTSAGLSLEEDQL 130
QY 569 PS-----SIKIINYKNISSSILLYNDP-----PEYPNGKITHYTYI 605
DB 131 PSGFPTIDMGPOLKVVKEGRATMLCAAGGNPDPEISWFKPLPVPDPAASNGRIKQLRSG 190
QY 606 AMELDTNRAPOIITINDSFLITGLKYYTKYKMRVAASHTDGBSSLSSEENDIFVRTSEDEP 665
DB 191 ALQIESSEE-----SDQGYE-----CVATNSAGTRYAPANLYVRVRVAP 232
QY 666 ESS--PODEV-----IDVTADIEIRLKWSPPEKPNGLIIAIEVLYKNIDTLYMKN 713
DB 233 RFSIPSSQEVMPGGSVNLTCVAGAMPYVKK-----MGAELTKEDEMPVGRN 283
QY 714 TSTTDIILNRLPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSA-----PENITYK 769
DB 284 VLELSNMRS-----ANYTCA-----ISSLGMIEATAQVTVKALPKPDIPLVT 328
QY 770 NISSGBIELFLPPSPNGIHKYTYLKR--SNGNEERTINTSITONIKVLKKTQYI 827
DB 329 ETTATSVTLTW--DSGNTEPVSYGQYRAAGTDGPFQEVGVASTRISYIGLSPESEYA 386
QY 828 IEVSASTLGEVRSAPISILTEEDAPDPPQFSVKQLSGVTVKLSWQPPLEPNGLIY 887
DB 387 FRVLAVNSIGRPPSEAVARTGEQAPSSPRVQARMLSASTMLVQWPEEPPEFNLVRG 446
QY 888 YTVY-----VNRSLKTIINTVETSL--ELSDLYNVEYSAYVTASTRFGDGKTG 935
DB 447 YRYVYTPDSRPLSAMKHEN-----TDAGLLTVGSLLPGVYISLRVLAFTAVGGPP- 499
QY 936 SNIISFOTEGADSPDKDYYANLSSSS--ILFWTPPSKPNGLIYQYVYVYNTSGTFM 994
DB 500 SPTIQVKTOQGVARSAD--FQANAESDTRIQLSWLLPQER-IVKYELVYV-----AAE 551
QY 995 QNFTEHLELNDFDMVTSTIIDKLTIESYTFWLTASTSVGNKSSDIEVYTDQDIPE 1054
DB 552 DEGOQKHVTFD--PTSSYTLDELKPDTLRYFQLAARSGLGVG-VFTPTVEAXTAQSTPS 607
QY 1055 GFVGNLTYESISTAINVSWPPA--OPNGLV-----FYVSLILOQTPRHVRPLVYTER 1108
DB 608 APPQKVTCTVSTGTVTVRVSVPADSRNGIITQYSVAEAVDGEKRVVDGISEHS 667
QY 1109 SIYFDNLEKYTDYILKITPSTKEGSDTYTAQYIKTEEDVPETSPINTFKNLSSTSVL 1168
DB 668 SWDLLGLEKWTEYRVWVRAHTDVG-PPGPSSVLVTRDEDVPSGPPRKVEVEPLNSTAVH 726
QY 1169 LSWDPPV--KPNCAIISYDLTLOGPNEYSFITSDNYIIEELSPTLISFFFAAATKRG 1226
DB 727 VSKFLVPNKKHQGQIRGYQT-----YVLENGEP----- 756
QY 1227 LGPSSILFFYTTDESVP LAPPNLTINCTSDFVWLKXWSPSLGGIVKVYFKIHEHTD 1286
DB 757 RQQPIL-----QDVMLEAQ----- 771
QY 1287 TIYKYNISGPKTBKLVGLFSPVSTYIRVSAPTKVGNQFNSVVKFTTQESVPDVVQNM 1346
DB 772 -----ETTISGLTPETTYISITVAAYTTKGDGAR-SKPKVVTITGAVPG-RPTM 817
QY 1347 QCMATSWQSVLKWDDPPKANGIITQMVTVR-----NSTKVSQDEMYTPIKLANT 1400
DB 818 MVSTAMHTALLQWHPKPELPGELLGVLRYQYRAREARPNTIDFGDDOCHFVTGLHKA 877
QY 1401 SYVPKVRASACESGDESTCHVSTLPETVPS-VPTNIAESDVOSTSATITWIRPDITL-- 1457
DB 878 TYVFLAAKNRAGBEPFEKBITT-PEDVPSGPPQNLRVTLGTSTTETLTWDDPP--VLAE 934

QY 1458 --GYFONY-----KITTLQRAQCKEWESECEVYQKIQYLYEAHLTEETVYGLKKFRWY 1510
DB 935 RRGKXITNYTVVVRDINSQLEQNVN-----DTHL-----TLGLKPDPTTY 975
QY 1511 RFQVAASTNAGYCNASNISTKTLPDPDPDENHVHATSPFSISISSEP-----AVI 1565
DB 976 DIKVRATSKGAGPLSPSISQSRTP-VEQFTYKFNFRVAAAMKTSVLLSHEVPDYSKAV- 1033
QY 1566 TQPTCYLIDVKSVYDNDDEFNISFKSNEENKTIIEKDLIFTRYSVVITAFPTNISAAYVE 1625
DB 1034 --PKILYNGQSVEVDGHSKRL-----IADLPQNTYESFVL--MNRGSSAGGLQ 1079
QY 1626 GKSSAEMIVTLESAP-----KDPNNMTFQKI--PDEVTKFQLTFLPPSPQNGNIQV 1676
DB 1080 HLVSINTAPDLLPQKPLPASAFIEDGRFSLSPQVQDPSLVRWFYIVVVPIDRVGNL-- 1137
QY 1677 YQALVYREDPTAVQIHNLSIIQKTNFTVIAMLEGKLGHTYNI SVAVNSAGACPKVPM 1736
DB 1138 --LAPRWTPPELEDEL-----LEAIEQSEBKQ----- 1164
QY 1737 RITMDIKAPARPKTPPIYDATGKLLVTSTTITIRMPICYSDDDHGPKNVQVLAETG 1796
DB 1165 -----RRRRROAERLKPVAAQVDVLPDTFTLGDK----- 1194
QY 1797 AQHNGVNTKYDAYFNKARPYPFTNEGFPNPPCTEKGTFSGNEEYIIGADNACMIPGNE 1856
DB 1195 -----KSYRGFYNR----- 1203
QY 1857 DKICNGPLPKKQY-----LPKFRATNIMGQFSDSDYSDPVKTLGEGLSRBTVEIILSVT- 1911
DB 1204 -----PLSPDLSYQCFVLASKEPMDQRYASSPYSDEIVVQVTPAQOQSEBEMLVWTG 1257
QY 1912 --LCILSIITLGTAFAPARIRQKQEGGTYPQDAEIIIDTKLKLQDLITVADLELKDER 1969
DB 1258 PVLAVILIILVAILLFRKR-----THSPSSKDEQSIGLK-DSLLAHSDDPVEMRR 1309
QY 1970 LT-----RPISKKSFLQHVHELCTNNLKFQEPSEL-----PKFLQDLSSTDADLPW 2017
DB 1310 LNYQTPGMRDHPPIPTITDLADNIERLKANDGLKFSQEYESIDPGQOFTWENSEV---- 1365
QY 2018 NPAQRFPNPKPVNNNRVKLIADASVPGSDYINASYISGLCPNFIATQGPLGTVGD 2077
DB 1366 NPKPKRIADVIAY-DHSRVLLTSIDGPGSDVINANTIDGKQKQNIATQGPLPETMGD 1424
QY 2078 FWRVWETRAKTLVMLTQCFEGRIRCHOYWPEDNKNPVTVPFGDIVITIKLMEDVQI-DWTI 2136
DB 1425 FWRVWEQRTATVMMTRLEKSRVKCDQYWPV--RGTYTYGLIQVT-LVDTVELATYTM 1481
QY 2137 RDLKIERHG--DCMVTROCNFTAWBHGVPENSAPLIHFVKLVRASRAHDTTPMIVHCSA 2194
DB 1482 RTFALHKSGSSEKRELRQPFMAWPDHGVPEYPTPIFLAFLRRVKACNPLDAGPMVHCSA 1541
QY 2195 GVGRTGVFTALDHLTOHINDHDFVDIYGLVAELSERMCMQVONLAQYIFLHCILDLN 2254
DB 1542 GVGRTGCFIVIDAMLERMKHKTVDIYGHVTCMSQRNTWVQTEDQYVFIHEALL-AM 1600
QY 2255 KGSNOPICFVNTYSALQKM-----DSLDAEGDVEL 2284
DB 1601 CGHTEVLARNLTAHQKLGQVPPGESVTAMELEFKL 1636
RESULT 9
Q64604
ID Q64604 PRELIMINARY; PRT; 1898 AA.
AC Q64604; Q63294; Q63295; Q63296;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Protein-tyrosine phosphatase, receptor-type, F polypeptide precursor
DE (EC 3.1.3.48) (LAR protein) (Leukocyte antigen related) (Leukocyte
DE common antigen-related phosphatase) (Protein-tyrosine-phosphatase)
DE (Phosphotyrosine phosphatase) (PTPASE).
GN Name=Lar;

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
 RX MEDLINE=94347119; PubMed=8068021;
 RA Zhang W.R., Hashimoto N., Ahmad F., Ding W., Goldstein B.J.;
 RT "Molecular cloning and expression of a unique receptor-like protein-
 tyrosine-phosphatase in the leucocyte-common-antigen-related phosphate
 family.";
 RL Biochem. J. 302:39-47(1994).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORMS PRLAR4.0 AND PRLAR631).
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HIPPOCAMPUS;
 RX MEDLINE=95146548; PubMed=7844155; DOI=10.1083/jcb.128.3.415;
 RA Zhang J.S., Longo F.M.;
 RT "LAR tyrosine phosphatase receptor: alternative splicing is
 preferential to the nervous system, coordinated with cell growth and
 generates novel isoforms containing extensive CAG repeats.";
 RL J. Cell Biol. 128:415-431(1995).
 RN (3)
 RP SEQUENCE OF 1035-1898 FROM N.A., AND MUTAGENESIS.
 RC TISSUE=HYPOTHALAMUS;
 RX MEDLINE=92011772; PubMed=1918076;
 RA Pot D.A., Woodford T.A., Remboutsika E., Haun R.S., Dixon J.E.;
 RT "Cloning, bacterial expression, purification, and characterization of
 the cytoplasmic domain of rat LAR, a receptor-like protein tyrosine
 phosphatase.";
 RL J. Biol. Chem. 266:19688-19696(1991).
 CC -1- FUNCTION: IT IS POSSIBLE THAT LAR IS A CELL ADHESION RECEPTOR. IT
 POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY
 (PTPASE). IT MODULATES SIGNALING BY THE INSULIN, EPIDERMAL GROWTH
 FACTOR, AND HEPATOCYTE GROWTH FACTOR TYROSINE KINASE RECEPTORS.
 MAY PLAY A ROLE IN NEURITE OUTGROWTH AND/OR CELL-CELL INTERACTIONS
 MODULATING SYNAPSE FORMATION. MAY INFLUENCE CELL MOTILITY.
 CC -1- FUNCTION: THE FIRST PTPASE DOMAIN HAS ENZYMATIC ACTIVITY, WHILE
 THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE
 FIRST ONE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN
 TYROSINE + PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=PRLAR;
 CC IsoId=Q64604-1; Sequence=Displayed;
 CC Name=PRLAR631;
 CC IsoId=Q64604-2; Sequence=VSP_050409, VSP_050410, VSP_050413,
 VSP_050415;
 CC Name=PRLAR4.0;
 CC IsoId=Q64604-3; Sequence=VSP_050415;
 CC Name=PRLAR18.1;
 CC IsoId=Q64604-4; Sequence=VSP_050414, VSP_050415;
 CC Name=PRLARCB9;
 CC IsoId=Q64604-5; Sequence=VSP_050411, VSP_050412;
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN NEURAL OR
 NEUROMUSCULAR TISSUE.
 CC -1- DEVELOPMENTAL STAGE: THE ALTERNATIVELY SPLICED ISOFORMS ARE
 DEVELOPMENTALLY REGULATED.
 CC -1- SIMILARITY: EXTRACELLULAR REGION TYPIC OF A CAM FAMILY (3 IG-LIKE
 DOMAINS FOLLOWED BY 8 FIBRONECTIN TYPE III-LIKE DOMAINS), AND A
 CYTOPLASMIC REGION COMPOSED OF TWO PTPASE DOMAINS (BY SIMILARITY).
 CC EMBL; L11586; AAC37655.1; -;
 CC EMBL; M60103; AAA41510.1; -;
 CC EMBL; X83505; CAA58495.1; -;
 CC EMBL; X83546; CAA58537.1; -;
 CC FIR; S46216; S46216.
 CC HSP; P10586; ILAR.
 CC GO; GO:0016021; C:integral to membrane; IEA.
 CC GO; GO:0016787; F:hydrolase activity; IEA.
 CC GO; GO:0005515; F:protein binding; IEA.

DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR003962; FNIII_subd.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_c2.
 DR InterPro; IPR001005; MYD_DNA_binding.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; Tyr_PP.
 DR Pfam; PF00041; fn3; 7.
 DR Pfam; PF00102; Y_phosphatase; 2.
 DR PRINTS; PR00014; ENTYPPELII.
 DR PRINTS; PR00700; PRTPHPHTASE.
 DR SMART; SM00060; FN3; 8.
 DR SMART; SM00408; IGc2; 2.
 DR SMART; SM00194; PTPC; 2.
 DR PROSITE; PS50853; FN3; 8.
 DR PROSITE; PS50835; IG_LIKE; 3.
 DR PROSITE; PS00037; MYE_1; UNKNOWN 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
 KW Alternative splicing; Cell adhesion; Glycoprotein; Hydrolase;
 FT Immunoglobulin domain; Receptor; Repeat; Signal; Transmembrane.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 1898 PROTEIN-TYROSINE PHOSPHATASE, RECEPTOR-
 TYPE, F POLYPEPTIDE.
 FT DOMAIN 28 1254 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 1165 1169 POLY-ARG.
 FT TRANSMEM 1255 1275 POTENTIAL.
 FT DOMAIN 1276 1898 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 47 114 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 149 214 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 246 305 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 317 ? FIBRONECTIN TYPE-III.
 FT DOMAIN 435 ? FIBRONECTIN TYPE-III.
 FT DOMAIN 1361 1607 PROTEIN-TYROSINE PHOSPHATASE.
 FT DOMAIN 1650 1898 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 1539 1539 BY SIMILARITY.
 FT ACT_SITE 1830 1830 BY SIMILARITY.
 FT DISULFID 54 107 POTENTIAL.
 FT DISULFID 156 207 POTENTIAL.
 FT CARBOHYD 253 298 POTENTIAL.
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 721 721 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 941 941 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 600 707 CTAQSTPAPQKVCSTGTVSTVSVWVPPADSRNGIIT
 QYSVAYEADVGDKRHHVVDGSRHSSMDLGLKWKTEYR
 VVWRATHDVGCPESPVLVRTDEDV -> RTAQM (in
 isoform PRLAR631).
 FT VARSPLIC 771 771 FTId=VSP_050409.
 FT VARSPLIC 771 771 Q -> QMRPSEEDY (in isoform PRLAR631).
 FT VARSPLIC 810 810 /FTId=VSP_050410.
 FT VARSPLIC 811 1898 A -> AGE (in isoform PRLARCB9).
 FT VARSPLIC 812 1004 Missing (in isoform PRLARCB9).
 FT VARSPLIC 1276 1276 /FTId=VSP_050412.
 FT VARSPLIC 1316 1316 Missing (in isoform PRLAR631).
 FT VARSPLIC 1539 1539 /FTId=VSP_050413.
 FT VARSPLIC 1777 1777 K -> KSKQE (in isoform PRLAR18.1).
 FT VARSPLIC 1073 1073 /FTId=VSP_050414.
 FT MUTAGEN 1539 G -> GSAPECPNISS (in isoform PRLAR4.0).
 FT CONFLICT 777 C->S: LOSS OF ACTIVITY.
 FT CONFLICT 1073 T -> S (IN REF. 2 AND 3).
 FT CONFLICT 1073 G -> S (IN REF. 2).

FT CONFLICT 1434 1434 I -> T (IN REF. 2 AND 3).
FT CONFLICT 1639 1639 G -> N (IN REF. 2 AND 3).
FT CONFLICT 1643 1644 RA -> HT (IN REF. 2 AND 3).
SQ SEQUENCE 1898 AA; 211492 MW; DF4D7E46F5896F4B CRC64;

Query Match
Best Local Similarity 9.9%; Score 1181; DB 2; Length 1898;
Matches 478; Conservative 273; Mismatches 704; Indels 518; Gaps. 72;

QY 461 PEIVNIVEMVGLYGEAEMSSDLHSLATFIYNHDPDKNFPARNRAEDQTSPPVT-----515
DB 33 PVPVKVPEQIGLGG-----VAFVC-----QATGPKPRIITWKKG 70

QY 516 --TRNOYITDIAEQLSVIRLVPF-----THMISVSAFTMGE--GPPTVLSVRTRQQV 568
DB 71 KKVSSORFVIEPDDGAGSVLRIOPLRVORDEAIEYCTATNSIGINTSAKLSVLEEDQL 130

QY 569 PS-----STKIINYKNISSSILLYWDP-----PEYPNGKITHYTIY 605
DB 131 PSGFPTDMGPQLKVVKEKARTATMLCAAGNPDPEISWFKDPLFVDPASNGRIKOLRSG 190

QY 606 AMELDNTRAFQIITIDNSPLITGLKKTYYKMRVAASTHDGSSLSSEENDIFVRTSEDEP 665
DB 191 ALQIESSE-----SDQKYE-----CVATNSAGTRYAPANLYVVRVAP 232

QY 666 EGS--PQDVEVIDTADIELKWSPEKPGNGIIIAEYEVLYKNIDTLMYKNSTTIDILRN 723
DB 233 RFSIPSSQEVN-----PGG-----NVNLTCVAVGA 258

QY 724 LRPHLYNIVSRYSYTRFGH--GNQVSSLLSVRTSETVSPDSAPENITYKNISS-----773
DB 259 PMPYVMMWGAEBLTKEDEMPVGRNVLSESNMRS-----ANYTCVAISSLGMEAT 310

QY 774 GRELSFLPSSPNGIHKYT--IYLKRSNGNEE-----RTINT-----T 811
DB 311 AQTVKALPKPDLVVTETATSVLTWDSGNTPEVPSYGIQYRAAGTDGPFQVGVGA 370

QY 812 SLTONIKVLKKTQYIIEVSASLTKGEVRSAPISILTEEDAPSDPPQDFSVKQLSGVTV 871
DB 371 STRYSIGGLSPFSEYAFVLAVNSIGKGPSEAVRARTGEQAPSSPRRVOARMLSASTM 430

QY 872 KLSQWPLEPENGIIYYTY-----VNRSSLKINTVETSL--ELSDLDYNYEY 919
DB 431 LVQWEPPEEPNGLVGRYVYTPDSSRRPLSAMKHN-----TDAGLLTVGSLLPGITY 484

QY 920 SAVYASTRGDCGTGNSIISFOTPEGAPSDPKOVVYANLSSSIILFWTPSKENGII 979
DB 485 SURVLAFTAVGDGPP--SPTIQVKTQGVPAQ--PADFOAKAESDTRIQLSWLLPPQBR--II 541

QY 980 QYYSVYRNSTSGTFMQNFTLHETLNDPDMNTVSTIIDKLITFSYTYFMTASTSVGNNGK 1039
DB 542 KYELVW-----AADEBGGQHKVTPD--PTSSVTLLEDKPDTLXHFQLAARSDLGVG-V 592

QY 1040 SSDIIEVYTDQDIPTEGFGVNLTVESISSAINVSWVPPA--QNGLV-----FYVSLIILQ 1093
DB 593 FTFVEACTAQSTPSAPPQKVTCSGSIITVRVSWVPPADRSNGIITQVSYVAEAVDGE 652

QY 1094 QTPHRVRPLVTVERSIYFNDLEKTYDILKITPSTEKGSFDYTAQLYKTEEDVPETS 1153
DB 653 DRKRHRVVDGISREHSSWDLGLEKTEYRVMVRAHTDVG--PGPESPVLVTRDDEVPSPG 711

QY 1154 PIINTFKNLSSTSVLSWDDPPV--KPNGATISYDLTLQGNENYSPTISDNYIILELSP 1211
DB 712 PRKVEVEPLNSTAVHVSXKLPVKNQHQGIRGQVT-----VYRLENGEP 756

QY 1212 FTLYSFPFAARTKGLGPSILFFYTDSEVPLAPPQNLTLINCTSDFWLKNWSPSPGPG 1271
DB 757 -----RGQPII-----QVWLAEAQ-----771

QY 1272 IVKYSGFKIHEHTDIYYKNISGFKTEAKLVGLEPVSITYSIRVSAFTKVGNGQPSNVV 1331
DB 772 -----ETTISGLTPTTYSITVAAYTTKGDGAR--SKPK 803

RESULT 10

Q6PG86

ID Q6PG86

AC Q6PG86;

PRELIMINARY;

PRT; 1529 AA.

QY 1332 KETTOESVDPDVONMOCMATSWQSVLVKWDPPKANGIITQYMTVVER-----NSTKVS 1385
DB 804 VVTTTGAVEG-RPTMMVSTAMHTALLQHPHPGELLGYRLQVRRDAEPNPTDFG 862

QY 1386 PQDHMYTTFIKLANTSYPFKVRASTAGSGDESBTCHVSTLPETVPS-VPTNIAFSDVQST 1444
DB 863 KDDQHTVTYGLHKGATYIFRLAAKNRAGPGBEFKEIIT--PEDAPSGFPQNLRTVGLTTS 921

QY 1445 SATLTWIRPDTIL-----GYFQNYKITTLQRAOKCKEWESEECVEYQIKIYLV-----1493
DB 922 TTLEAWDPP--VLAERNGRITNYTVV-----YRDINSQHELQNVTVGD 961

QY 1494 AHLTEETVYGLKFKMYREFQVAASNAGYGNASNMISTKTLPGPPDPENHVHVAATSPF 1553
DB 962 VHL-----TLGLKPDOTYDIKVRHRTSKAGPLSPSIQSRTPM-MEQVAFKFRVAAMKT 1017

QY 1554 STISWSEF-----AVITGTCYLDIVKSVDMDEFNIFSIKSEENKTIIEIKOLEIFTRY 1608
DB 1018 SVLLSWEVSDSYKSAV--PFKILYNGQSVEDVGHSMRKL-----IADLQNTY 1064

QY 1609 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAP-----KOPPPNMTFOKI--PDEVTK 1659
DB 1065 SPVL--MNRGTSAGGLQHLVSRITAPDLLPQKPLPASAFIEDGRFSLSPQVQDPSLVRW 1122

QY 1660 FQITFLPPSQPNQNTQVYQALVYREDDPTAVQIHNLSIIQKNTTFFVIAMLEGLKGHTYN 1719
DB 1123 FYVWVVIDRVGNL-----LAPRWSTPBELEDEL-----LEAIEGEEKQ 1164

QY 1720 ISVAVNSAGAGPKVPMRITMDIKAPARKTPTIYDATGKLLVTSTTITRMPICYYS 1779
DB 1165 -----RRRROAERLKPVAQVDELPTFTL-----1191

QY 1780 DDHGPIKNQVQLATETGAQHDGNVTWKYDAYFNKARPYFTNEGFPNPPCTEGTKFSGNE 1839
DB 1192 ---GDCKN-----YRGFYNR-----1203

QY 1840 EYIIGADNACMIPGNEDKICNGPLKPKQY-----LFKFRATNIMQFTDSYSDPVKTL 1895
DB 1204 -----PLSPDLSYOCFVLASKEPMDQKRYASSPSDIIVVQ 1240

QY 1896 GEGLSERTVEIILSVT--LCILSIILGLTAFAPARIROKQEGCTYSPQDAEIIDTKL 1952
DB 1241 VTPAQOOEPEPEMLWVTGPVLAVILIIILVIALLLFKRK-----THSPSSKDBOSIGL 1293

QY 1953 KLDQILITVADLELXDBERLT-----RPISKKSPLOHVEELCTNNNLKFOEFSEL--2001
DB 1294 K-DSLLAHSSDPFVEMRLNYQTPGMRDHPPIPIITDLADNIERLKANDGLKFSQEYSDIP 1352

QY 2002 -PKFLODLSSTADADLPWNRKARFPNPKYNNNNRVKLITADASVPGSDVINAYISGILC 2060
DB 1353 GQOFTWENSSEV---NKPKNRYANVIAY--DHSRVLLTSIDGVPGSVDYINANYIDGYSK 1407

QY 2061 PNEFIATQGLPCTGDFWRWYWEIATKVLMTQCFEGRIRCHOYWPEDNKPVTVFGD 2120
DB 1408 QNAYIATQGLPETGDFWRWYWEIATKVLMTQCFEGRIRCHOYWPEDNKPVTVFGD 2120

QY 2121 IVITTKLMEVDQI--DMTIRDLKIERHG--DCMTVROCNFTAWPEHGVHPENSAPLIHFVKLV 2177
DB 1466 IQVT-LVDVTELATYMTTFALHKSSESSEKRELROQFVWPDHGVPEYPTPLAFLRV 1524

QY 2178 RASRAHDTTPMIVHCSAGVGRGTGFVIALDHLTQHINHDHPVDIYIGLVAEILSRMCMQVN 2237
DB 1525 KACNPLDAGPMVHCSAGVGRGTGCFIVIDAMLERMKHEKTVDIYIGHVTCMRSQRNTMVOT 1584

QY 2238 LAQYFLHQCIIDLLSNKGSQPICFVNTYSAQKM-----DSLDAEGDVEL 2284
DB 1585 EDQYVPIHEALLE--AAMCGHTEVLARNLYAHTIQLKQGVPPGSSVTAMELEFKL 1636

Db 1090 DVLKTKPFVIGKNTSDGMITVELPEVLVNDKINGYIVIVPLKSRGK-----FIKPWE 1143
Qy 1686 DPTAVQIHLN--SIIOKTNFTVIALEGLKGGHTYINISYVAVNSAGAGKVPWRITMDLK 1743
Db 1144 SPDEMEFDELLKDIRK-----RSLRPR 1167
Qy 1744 APARPKTKPTFYDATGKLLVTSTTITIRMPICYSDDHGPIKNVQVLATETGAGHDGNV 1803
Db 1168 REAEQK---PYIAHFLLPTEFTL-----GDQ----- 1192
Qy 1804 TKWYDAYFNKARPYFTNEGFPNPPCTEGTKPSGNEBIVIIIGADNACMPGNEDKICNGP 1863
Db 1193 -KQYGGF-----ENKQLONGQYVEFVLA---VIEHSES----- 1222
Qy 1864 LXPKKOYLKFRATNIMGQTFDSYSDPKVTLEGLSERVE-----ILSVTLCLISII 1918
Db 1223 -----AMFATSPYSDPVVSMEIDPQMTDBEEGLIWVGPVLAVVFII 1265
Qy 1919 LLGTALFAFARIRQK-----OKEGTYSPQD-ABIIDTKLKDOLIT---VADLE 1964
Db 1266 CIVIALLLYRKRTESDSRKSSLPNSKEIPSHNPTDPVELRLNFQTPGMANHPPIPILE 1325
Qy 1965 LKDERLTRPISKSFQHVHELTNNLNKLPQEFSELPKFLQDLGSTDADLPWNAKRNK 2024
Db 1326 LED-----HIERLKANDNLKFSQYESIDPG-QQFTWEHSNLEVNKPKRY 1370
Qy 2025 PNKPYNNNRVKLIADSVGSDYINASYISGLYCPNEFIATQGLPCTGVDGFRMWWE 2084
Db 1371 ANVIAY-DHSRVLLSAIDIGPSDYINSYIDYRKQNAIYATQGPLPETFGDFWRMWWE 1429
Qy 2085 TRAKTLMVLTQCEKGRIRCHOVWPNEDKPVTVFGDIVITKLMDVOI-DWTIRDLKIBR 2143
Db 1430 QRSATVMMTKMEERSIKCDQWP--SRGTETYGILQVT-LDVTVELATYTVTFALYK 1486
Qy 2144 HG--DCMTVRQCNFTAWPEHGVNPENSAPLIHFVKLVASRAHDTTTPMIVHCSAGVGTGV 2201
Db 1487 NGSSEKREVRFQFTAMPDGHVPEHTPFLAFLRVKTCNCPDAGPMVHCSAGVGTGC 1546
Qy 2202 FTALDHLTQINDHDVFDIYGLVAELRSEKCMQVNLQAYIFLHCILDLNKSQNP 2261
Db 1547 FNVIDAMLERIRHEKTVDIYGHVTLMAQRNMYVQTEDQYIFTHDALLEAVTCGNTVEPA 1606
Qy 2262 CFNYSALOKMDSLDAMEGDVLEWE 2287
Db 1607 RNL-YAIOKLTQIEPCENVTGMELE 1631

RESULT 12

Q9QW67
ID Q9QW67 PRELIMINARY; PRT; 1887 AA.
AC Q9QW67;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE LAR, leukocyte common antigen-related PROTEIN-TRANSMEMBRANE receptor
DE Phosphotyrosine phosphatase.
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92278755; PubMed=1317540;
RA Yu Q., Lenardo T., Weinberg R.A.;
RT "The N-terminal and C-terminal domains of a receptor tyrosine
RT phosphatase are associated by non-covalent linkage";
RL Oncogene 7:1051-1057(1992).
DR HSP3; P10586; ILAR.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN_III.

DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR001005; MyD_DNA_binding.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_PP.
DR Pfam; PF00041; fn3; 7.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PRO0014; ENTYPETIII.
DR PRINTS; PRO00700; PRYPHPTASE.
DR SMART; SM00060; FN3; 8.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS50853; FN3; 8.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00037; MYB_1; UNKNOWN 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00566; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
SQ SEQUENCE 1887 AA; 210453 MW; B84B33E7E4E70281 CRC64;

Query Match 9.7%; Score 1162.5; DB 2; Length 1887;
Best Local Similarity 24.3%; Pred. No. 9e-52;
Matches 476; Conservative 284; Mismatches 709; Indels 489; Gaps 74;

Qy 461 PEINIVPMVCLYEGSAEMSDLSLATFIYNHSHDPKFNFPARNABQTSPPVVT----- 515
Db 23 PVFVKVBDQTLGG-----VASFC-----QATGEPRPRTIWMKKG 60

Qy 516 --TRNQYITDAEQLSVYVRLVPF---TEHMISVSFAFTIMGE-GPPTVLSVTRQOV 568
Db 61 KKVQSQRVEVIEFDGAGSVLRIQLRVQRDEAIECTATNSLGEINTSAKLSVLEEDQL 120

Qy 569 PSIKIIN-----YKNISSSIL-----LYM-----DPPEYNGKITHYTI 604
Db 121 PSLGLPTIDMGPKLVKVKARTATMLCAAGGNPDPEISWFKDPLPVPDAS-SNGRIKQLRS 179

Qy 605 YAMELDTNRAFOITIDNSFLITGLKTKYKMRVAASHTDGESESSIBENDIFVRTSEDE 664
Db 180 GALQIESSEE-----SDQKYE---CVATNSAGTRYSPANLIYVRRVA 221

Qy 665 PESS--PDQVEV-----IDVTABEIRLKSPPKPKNGIILIAEVLVYKNIDTYMK 712
Db 222 PRFSIPPSQEVMPGCVNLTCAVGAAMPYVKM-----MWGAELTDEKEMPVS 272

Qy 713 NTSSTTDIILNLRPHLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDAS-----PENITY 768
Db 273 NVLELSNVMS-----ANYTCA-----ISSGLMIEATAQVTVKALPKPPTDLVV 317

Qy 769 KNISSEIELSPLPPSSNGIILKTYIYIKR--SNGNEERTINTTSLTONIKVLKXYTOY 826
Db 318 TETTATSVLTW--DSGNTEPVSYGIQYRAAGTDGPPFQVDGVASTRYSIGLSPFSY 375

Qy 827 LIEVASLTKGRVRSAPISILTEBDAPDSPQDFSVKQSGVTVKLSQPPLENGIIL 886
Db 376 AFRVLAVNSIGKGPSEAVRARTGQAPSSPPRRVQARMLASATMLVQWDPB-ENGLVYR 434

Qy 887 YTYTV-----VNNRSKLTINVTETSL--ELSDLDYNNVYSAYVYASTRFGDKGT 934
Db 435 GYRVVYTPDSRRRPLSAWHKN-----TDAGLTTVGSLLPGITYSLRLAFTAVGDGPP 488

Qy 935 GNNIISFQTPPEGAPSDPKDQVYVYANLSSSIIILFWTPPSKPKNGIILQIYVSVYVNTSGTFM 994
Db 489 -SPTIQVTKTQQGVPAQAPDQAKAELDTRIQLSWLLPQGE--RIIKYEVYVW-----AAE 540

Qy 995 QNFTLHELNDPDMNTVSTIIDKLATFSYTFEWTASTSVGNKNSSDIIIEVYTDQDIDE 1054
Db 541 DEGOQHKVTFD---PTSSYTLDELXPDITLYHQLAARSDLVGV-VFTPTVEARTAGSTS 596

Qy 1055 GFVGNLTVYESISSTAINVSWPPA--QPNGLV-----FYTVSLIQOTPHVAPPLVTYER 1108
Db 597 APPQKVTCTVGSTTVRVSWVPPADSRNGIITQYSVAYEAVDGEDRKRHVVDGISREHS 656


```
Db 1056 SYEPV-----NATQKIDGLVGNHYIFRIQKASLGALGVAERHEHQTPIIAPPVPEPS 1110
QY 1752 PTPYDAGTKLLVTTITIRMPICYSDDHGHPIKQV-VLAPETGAQHDG-NVTKYWD- 1808
Db 1111 VTPL-----EVSRTSTIBSPROGYSFNAHGMVRSYVITIIABDVGNASGLEMPQSDV 1165
QY 1809 -AY-----FNKARPYTNEGFPNPPCT-EKQTFSGNBEIYIIGADNACWIPGNEDK--- 1858
Db 1166 QAYTWLWLPYQAIPEY-----NPFITSGSRKSLAEAEHFTGTANC-----DKHOA 1211
QY 1859 -ICNGPLKPKQYLFKFRATNIMGQTFSDSYDPVTKIGELSGTVEIILSVTLCLIST 1917
Db 1212 GYCNGLERAGTTRIKIRAFTEDEKFTDVIYSSPITT-----ERSDTVIVAAT---VSA 1262
QY 1918 ILGTAIFAPARIRQKQEGGTYSPQDAEIIDFKLX-----DGLIIVADLELKDRLTRP 1973
Db 1263 VLLVAMVLVV-----VYQHRCLQIRREASKLARMQDELAALPGSYITP---NRP 1308
QY 1974 ISKKSFLQVHELCTNNLKFQEPSELKFLQDLSSDADLPWNRKARFPNPKYNNN 2033
Db 1309 VHYKDFSEHYRIMSADSRFSEFEELKHVGRDQACSFANLPNRPKNRFTNILPY-DH 1367
QY 2034 NRVLKLADASVSGSDYINASYISGLYCPNEFIATQGLPCTVCGDFWRVWETRAKTLVNL 2093
Db 1368 SRFLQPVDDGSDYINANYMFGHNSPREFIVTQGLHSTREBFWRMCWESNRAIVML 1427
QY 2094 TOCFEGRIRCHOYVEDNKNPVTVFGDIIVTKLMEDVDQIDWTRDLKIRHGDGCMTVROC 2153
Db 1428 TRCFEGRKCDQYVPD-RVAMPYGDIKVQLIIDHYHDWSISEFVWERNCESRIMRHF 1486
QY 2154 NFTAWEHGVENSAPLHIFVKLVASRAHDTTPMTVHCSAGVGRGVFIADHLHTQHN 2213
Db 1487 HFTWPDGFPVPEPQSLVFRVAFRDVIGTDMRPIIVHCSAGVGRSGTFIADRLILQIH 1546
QY 2214 DHDFVDYIGVLAELSERMCWQNLQAYIFLHCQIIDLILSNK 2255
Db 1547 KSDYVDIFGIVFAMRKERVFMVQTEQQYVCIHQCLLAVLEGK 1588

RESULT 14
Q9W4F5
ID Q9W4F5 PRELIMINARY; PRT; 1767 AA.
AC Q9W4F5;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE CG6899-PA.
GN Name=Ptp4E; ORFNames=CG6899;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

```
RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupeki M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tuzy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, AE003432; AAF45998.1; -.
DR PIR, A49502; A49502.
DR PIR, B49502; B49502.
DR HSSP; P10586; 1LAR.
DR IntAct; Q9W4F5; -.
DR FlyBase; FBgn004368; Ptp4E.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
```


Qy	568	-VPSSIKIINYKNI-----SSSILYWPDPPEYPNGKITHYTIYAMELDTNRA	614
Db	304	GVIEALAIQITVKALPKPGTPVVTTESTATSIITLWDSGN--PGFVSYYIIQHKPKNSEEP	361
Qy	615	FQITTTD---NSFLITGLKKYTKYKRWVAASTHDGESSLSSEENDLFVRTSEDEPSSPQ	670
Db	362	YK--EIDGIIATTRYSVAGLSPGDYEFVRVAVANNIGRPASE--PVLTOTSETQTPSAPAR	417
Qy	671	DVEIVIDTADEIRLKWSPPEKPNGIIIAYEVLVKNIIDTLYMKNTSTTDI-----ILRL	724
Db	418	DVQARMLSSITILVQWKEPEEPNGIOGYRVYVTMDPTQHVNNMKHNVADSOITIGNL	477
Qy	725	RPHLTYNISVRSYTRFGHGNQVSSLSVRTSETVPDSAPENITYKNISGEIEBLSFLPDS	784
Db	478	VPOKTVSVKVLAPTSIGDG-PLSSDIQVITQTVGP-QGPLNFKFAEPESSETSILLSWTPPR	535
Qy	785	SPNGIIKKYTIYLKRSN-GNEERTINTTSITQNIKVLKKYQYIIIEVSASTLKGVRSA	843
Db	536	SDT--IASYELVORDQGEQRIITPEGTSYRLQGLKPNLSLYFRLSATSPOGLGASTA	593
Qy	844	PISILTEEDAPDPPQDFSVKQLSGVTVKLSWOPPLEPNPNGIILYYTYVWNRSLSKTINV	903
Db	594	EISARTWQ-KPSAPPQDISCTSPSSITILVSWQPP-----627	
Qy	904	TETSLEUSDLDYNNVEYSAYTASTRFQDGKTGSIISFQTPPEGAPDPPKDVVYANLSSS	963
Db	628	-----627	
Qy	964	SIILFWTPPSKPNGLIOYYSVYRNTSGTFMQNFTLHELNDPDMTVSTIIDKLTFPSY	1023
Db	628	-----PVKQNGIITFYSLKAAVDG---EDYKHEIICNGSD-TTKYLLQLQEKWTE	676
Qy	1024	YTFWLTAHSVGNKNSDIIIEVYTQDIDPEGFVGNLTYESISSTAINVSW---VPPAQP	1080
Db	677	YRITVTAHTDVGPWPESLSVL-IRTDSDVPSPRKEVEAVNATAVKVSWRSPVENKQH	735
Qy	1081	NLGVFYVYVLSILOQT--PRH--VRPPLVTVYRSIYFDNLEKYTDYILKITPSTFKGSDT	1136
Db	736	GQIRGYQVHVYKMBNGSPKSAIMLKDVMLADAQMIISGLQPETSYSLTVAYTTKGDGAR	795
Qy	1137	YTAQLYIKTEEDVPETSPINTFNKLSSTSVLLSWDPPVPKPGCALIISYDLTLQGNENYS	1196
Db	796	SKPKL-VSTTGSVPKGRPLV--INHQTOMTALIQWHPVDVTFG-----LQYRLKFG	844
Qy	1197	FITSNDYIIIEELSPFTLYSFFAAARTRKGLGPSSILFFYTDSDVSLAPPQNTLINCTS	1256
Db	845	-----RKDMPLTTLFSEKE-----860	
Qy	1257	DFVWLKWSPLPGGIIVKVSFKIHEHTDTIYYKNISGFKEAKLVGLPVPVTSYIRVS	1316
Db	861	-----DHFTATDIIHKGAS-----YVFRLS	879
Qy	1317	AFTKVGNGNOFNSNVKFTTOESVP-DVVQNMQCMATSWQSQVLVKWQPPKKA--NGIITQY	1373
Db	880	ARNKVGFGEEM--VKEISVPEEPTGPQNLHSEGTSTSVQLSWQPPVLAERNGVITKY	937
Qy	1374	MVTVERNSTKVSPODHM-----YTFIKLLANTSYVVKVRASTAGSGDES-TCHVSTL	1425
Db	938	TLLYRDINVPLLPMEHLIVPADTSMTILTGLKSDTTVDVKVRAHTSKRPGAYSPOFRTL	997
Qy	1426	PETVPSVPTNIAFSDQVSTSATITWIRPDILGVFQNY-----KITTLRAQCKEWES	1479
Db	998	PVQDSMFAPKPHVKAVMKTISVLLSWEIP-----ENYNPAILSKPFMMQDGMKVEE---	1047
Qy	1480	EECVEYOKIQVLYEAHYTEETVYGLKXFRYRQVAASTNAG--YGNASNWIISTKLPGP	1537
Db	1048	-----VDGRATQKLIYNLKEKYSVP---ALTNRGNSAGGLQHRVTAKTAP--	1090
Qy	1538	PDGPPENHVATSPFSSISISWSBPAVITGPTCYLIDVKSVDNDEFNISI-FKSNENK-	1595
Db	1091	-----DVLRTKP-----AFIGKTNLDGMI	1109

Qy	1596	TIEIKOL---EIFFRYSVWITAFGNISAAVYEGKSSAEMIVITLESAPKDPNNMTFQK	1652
Dy	1110	TVOLPDVPANENIKGYIIIIIVPLK-----KSRGKFI-----KPWESPEDELMDE	1153
Qy	1653	IPDEBTKFOLTFLPPSPQNGNIQVYQALVYREDDPATAVQIHNLSIIQKTWTFVIAMLEGL	1712
Dy	1154	LLKEISKR-----	1162
Qy	1713	KGGHTNYSVYVNSAGAGPKVPMRITMDIKAPARPKTRPTPIYDATGKLLVSTTTIIR	1772
Dy	1163	-----RSIRYGREVELK-----PYIAAHFDVLPTEFTL---	1190
Qy	1773	MPICYYSDDHGPIKNVQVLATETGAQHDGNVTWKYDAYFNKARPYFTNGBFPNPPCTEGK	1832
Dy	1191	-----GDD-----KHG-----GFTN-----K	1202
Qy	1833	TKFSGNEIYIIGA--DNACWIPGNEDKICNGPLKPKQYLFKFRATNINGQFTDSYSD	1890
Dy	1203	QLOGGQGYVFFVLAVMDHA-----ESKM-----YATSPYS	1233
Qy	1891	PVKTLG-----EGLSERTVEIITSLVTLCTILSIIILGTAFARIRQK-----	1933
Dy	1234	PVYSMDLDPQPIITDEESGLI-----WVVGPFVAVFIIICIVIAIIILYKRKAESDRKSS	1288
Qy	1934	---QKGGTVSPQDAEIIDTKLKDQLIITVADIELEKDERUTRPISKSKSFLOHVEELCTNN	1990
Dy	1289	LPNSKEVPSHHPTD-----PVELRLT-----NFQTPGMASHPPPIILELADHIERLKAND	1338
Qy	1991	NLKQOEFSFELPKFLODLSSTDADLPNRAKNRFPNIPKPNNNNRVKLIADASVPGSDYI	2050
Dy	1339	NLKPSQYESIDPG-QQFTWEHNSLEVNKPNRYANVIAI-DHSRVLLSIAIEGIPGSDYV	1396
Qy	2051	NASYISYGLCPNBFIAIQGPLPGTVGDFWFMVWETRAKTLVMLTQCPEKGRIRCHQWPE	2110
Dy	1397	NANYIDGRQKNAYIAIQGSLPETFGDFWMIWEQEA-TVVMTKLEERSRVKCDQWYP-	1454
Qy	2111	DNKPVTVFGDVIIVTKLMEDVOIDWTIHDLKIERHG--DCMTVROCNTANPEHGVBPNSA	2168
Dy	1455	---SRGTETHGLVQVT-LLDTVLHAILCPDICTLANGSSEKRVQFQFTAMPDGHGVPEHPT	1512
Qy	2169	PLIHFKVLVRASRAH--DTPPMIVHCSAGVGRGTGFIALOHLTQHINDHDFVDIYGLVAE	2256
Dy	1513	PVPSF--LTESQNLHPDAGPMVHCSAGVGRGTGCFIVIDAMLERIKHEKTVDIYGHVTL	1570
Qy	2227	LRSECMQVONLAQYIIFLHCICILLNKGNSNOPICFVNYSALQKMDSLDAMEGDVELEW	2286
Dy	1571	MRAGQNMVQTEQOYIIFHDALLEAVTCGTEVPARNL-YAYIQKLQIETGENVTGMEL	1629
Qy	2287	E	2287
Dy	1630	E	1630

Search completed: June 30, 2005, 15:04:22
Job time : 308 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:04:38 ; Search time 241 Seconds
(without alignments)
3667.457 Million cell updates/sec

Title: US-10-673-885-2
Perfect score: 11985
Sequence: 1 MDPLIFLLFGTSETQVD.....MDSLAMEGDVLEWEETM 2291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pcp.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pcp.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11985	100.0	2291	10 US-09-822-871-2	Sequence 2, Appli
2	11985	100.0	2291	15 US-10-673-885-2	Sequence 2, Appli
3	11807.5	98.5	2300	15 US-10-087-684-10	Sequence 10, Appl
4	11807.5	98.5	2300	15 US-10-218-779-10	Sequence 10, Appl
5	11797	98.4	2299	15 US-10-466-759-2	Sequence 2, Appli
6	11191	93.4	2281	15 US-10-087-684-6	Sequence 6, Appli
7	11191	93.4	2281	15 US-10-218-779-6	Sequence 6, Appli
8	10536	87.9	2301	10 US-09-822-871-4	Sequence 4, Appli
9	10536	87.9	2301	15 US-10-673-885-4	Sequence 4, Appli
10	10536	87.9	2302	15 US-10-087-684-37	Sequence 37, Appl
11	10536	87.9	2302	15 US-10-218-779-37	Sequence 37, Appl

12	4535.5	37.8	898	14	US-10-314-232-22	Sequence 22, Appli
13	4488	37.4	855	15	US-10-087-684-8	Sequence 8, Appli
14	4488	37.4	855	15	US-10-218-779-8	Sequence 8, Appli
15	1902.5	15.9	401	14	US-10-314-232-15	Sequence 15, Appl
16	1713.5	14.3	322	14	US-10-314-232-11	Sequence 11, Appl
17	1538.5	12.8	289	14	US-10-314-232-47	Sequence 47, Appl
18	1327	11.1	2037	15	US-10-087-684-39	Sequence 39, Appl
19	1327	11.1	2037	15	US-10-218-779-39	Sequence 39, Appl
20	1319	11.0	2029	15	US-10-087-684-38	Sequence 38, Appl
21	1319	11.0	2029	15	US-10-218-779-38	Sequence 38, Appl
22	1194.5	10.0	334	14	US-10-314-232-7	Sequence 7, Appli
23	1188	9.9	1912	17	US-10-772-636-64	Sequence 64, Appl
24	1160	9.7	1767	15	US-10-087-684-40	Sequence 40, Appl
25	1160	9.7	1767	15	US-10-218-779-40	Sequence 40, Appl
26	1158.5	9.7	1907	15	US-10-291-265-250	Sequence 250, App
27	1155.5	9.6	1897	17	US-10-482-029-52	Sequence 52, Appl
28	1155.5	9.6	1907	18	US-10-712-892A-34	Sequence 34, Appl
29	1150	9.6	1767	15	US-10-087-684-41	Sequence 41, Appl
30	1150	9.6	1767	15	US-10-218-779-41	Sequence 41, Appl
31	1141.5	9.5	1948	9	US-09-808-602-55	Sequence 55, Appl
32	1141.5	9.5	1948	10	US-09-800-198-45	Sequence 45, Appl
33	1109	9.3	1997	10	US-09-909-567B-54	Sequence 54, Appl
34	1108	9.2	1997	16	US-10-408-765A-2135	Sequence 2135, Ap
35	1108	9.2	1997	17	US-10-497-692-4	Sequence 4, Appli
36	1107	9.2	1997	15	US-10-634-027-2	Sequence 2, Appli
37	1026	8.6	1447	17	US-10-497-692-13	Sequence 13, Appl
38	966.5	8.1	1450	17	US-10-497-692-14	Sequence 14, Appl
39	942	7.9	176	14	US-10-314-232-9	Sequence 9, Appli
40	863	7.2	1238	15	US-10-366-547-7	Sequence 47, Appl
41	855	7.1	1367	15	US-10-369-493-5508	Sequence 5508, Ap
42	855	7.1	1367	15	US-10-369-493-5509	Sequence 5509, Ap
43	851.5	7.1	2041	13	US-10-087-192-1566	Sequence 1566, Ap
44	839.5	7.0	1502	9	US-09-808-602-54	Sequence 54, Appl
45	839.5	7.0	1502	10	US-09-800-198-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-822-871-2
; Sequence 2, Application US/09822871
; Publication No. US20030099942A1
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO01219
; CURRENT APPLICATION NUMBER: US/09/822,871
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2291
; TYPE: PRT
; ORGANISM: Human
US-09-822-871-2

Query Match 100.0%; Score 11985; DB 10; Length 2291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MDPLIFLLFGTSETQVDVSNVVGTRYDITISSITTYTSPVTRIVTPNVTKGPPV	60
Db	1	MDPLIFLLFGTSETQVDVSNVVGTRYDITISSITTYTSPVTRIVTPNVTKGPPV	60
QY	61	FLAGERVGSAGILLSNWTPPNPGRRIISIVYKVCPPMQVTYTVQVRKPSLEVLTN	120
Db	61	FLAGERVGSAGILLSNWTPPNPGRRIISIVYKVCPPMQVTYTVQVRKPSLEVLTN	120
QY	121	INPGTTYEIKVAENSAGIGVSDPFLPQTASAPGVNLTVEYANASAVKLITWLPQ	180

RESULT 2
US-10-673-885-2
; Sequence 2, Application US/10673885
; Publication No. US20040081644A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219DIV
; CURRENT APPLICATION NUMBER: US/10/673,885
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 09/822,871
; PRIOR FILING DATE: 04-02-2001
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2291
; TYPE: PRT
; ORGANISM: Human
US-10-673-885-2

Query Match 100.0%; Score 11985; DB 15; Length 2291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPLLIPLLLFIGTSETQDVSNVPCGTRVDITIISSITSTYTSPTVTRIVTPNVTKGPPV 60
DB 1 MDPLLIPLLLFIGTSETQDVSNVPCGTRVDITIISSITSTYTSPTVTRIVTPNVTKGPPV 60

QY 61 FLAGERVSGAGILLSNWTPNPNNGRIISYIVKVEKVCPMQVTVTVQVRKSPDSLEVLTN 120
DB 61 FLAGERVSGAGILLSNWTPNPNNGRIISYIVKVEKVCPMQVTVTVQVRKSPDSLEVLTN 120

QY 121 LNPCTTYEIKVAENSGAGVFDPLQTAESAPGKVNLTVEAYNASAVKLIWYLPKQ 180
DB 121 LNPCTTYEIKVAENSGAGVFDPLQTAESAPGKVNLTVEAYNASAVKLIWYLPKQ 180

QY 181 PCKGITSFKLSVKHARSIGVVDVSRVEDILTGKLPCECNENSESFLWSTASPSPTLGRV 240
DB 181 PCKGITSFKLSVKHARSIGVVDVSRVEDILTGKLPCECNENSESFLWSTASPSPTLGRV 240

QY 241 TPSSRTTHSSSTLTQNEISSVWKEPISFVVTHLRPVTTYLFEVSAATTEAGYIDSTIVRT 300
DB 241 TPSSRTTHSSSTLTQNEISSVWKEPISFVVTHLRPVTTYLFEVSAATTEAGYIDSTIVRT 300

QY 301 PESVPGPPQNCVTGNITGKSFILMDPPTIVTGKFSYRVELYGPGSRILDNSTDKLKA 360
DB 301 PESVPGPPQNCVTGNITGKSFILMDPPTIVTGKFSYRVELYGPGSRILDNSTDKLKA 360

QY 361 FTNLTPTTMDVYIAAETSAGTGPKSNISVFTPPDVPDVGAVFDLQLAEVSTQVRIITWKKP 420
DB 361 FTNLTPTTMDVYIAAETSAGTGPKSNISVFTPPDVPDVGAVFDLQLAEVSTQVRIITWKKP 420

QY 421 RQPNGIINQYRVKVLVPETGIIILENTLLTGNNYEIINDPMAPEIVNIIVPEPMVGLYEGSAEM 480
DB 421 RQPNGIINQYRVKVLVPETGIIILENTLLTGNNYEIINDPMAPEIVNIIVPEPMVGLYEGSAEM 480

QY 481 SSDLSLATPIYNHSDPKPNPARRAEDQTSPTVTRNQVITDIAEQLSYVIRRLVPPT 540
DB 481 SSDLSLATPIYNHSDPKPNPARRAEDQTSPTVTRNQVITDIAEQLSYVIRRLVPPT 540

QY 541 EHMISVSFTIMEGEGPTVLVSTROQVPSIKIINVKNISSSIIILYNDPPPYNGKIT 600
DB 541 EHMISVSFTIMEGEGPTVLVSTROQVPSIKIINVKNISSSIIILYNDPPPYNGKIT 600

QY 601 HYTIYAMELDNRAFOITIDNSFLITGLKKYTKYKMRVAAS*HDSSESLSENDIFVRT 660
DB 601 HYTIYAMELDNRAFOITIDNSFLITGLKKYTKYKMRVAAS*HDSSESLSENDIFVRT 660

QY 661 SEDEPSSQDVVEIDVTADIEIRLKWSPPEKNGIIIAVEVLKNIIDTLYMKWTSITDII 720
DB 661 SEDEPSSQDVVEIDVTADIEIRLKWSPPEKNGIIIAVEVLKNIIDTLYMKWTSITDII 720

661 SEDEPSSQDVVEIDVTADIEIRLKWSPPEKNGIIIAVEVLKNIIDTLYMKWTSITDII 720
721 LRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSDAPENITYKNISSEIELSP 780
721 LRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSDAPENITYKNISSEIELSP 780
781 LPSSPENGIIKKYTIIVLKESNGNEERTINTTSITQNIKVLKKYQTVIIEVSASTLKGEV 840
781 LPSSPENGIIKKYTIIVLKESNGNEERTINTTSITQNIKVLKKYQTVIIEVSASTLKGEV 840
841 RSAPISILTEEDAPSDPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSLKT 900
841 RSAPISILTEEDAPSDPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVVWNRSLKT 900
901 INVTETSLBLSLDLYNVVYSAYVTASTRFGDGKGTGSIISFQPEGAPSDPPKDVVYANL 960
901 INVTETSLBLSLDLYNVVYSAYVTASTRFGDGKGTGSIISFQPEGAPSDPPKDVVYANL 960
961 SSSSIILFWTPPSKNGIIQYYSVYRNTSGTFMQNFTLHFLTNDNDNMTVSTIIDKLTII 1020
961 SSSSIILFWTPPSKNGIIQYYSVYRNTSGTFMQNFTLHFLTNDNDNMTVSTIIDKLTII 1020
1021 FSYTFLWTASTSVGNGKSSDIIIEVYTDQDIPGFGVGNLTYESISSTAINVSWVPPAQP 1080
1021 FSYTFLWTASTSVGNGKSSDIIIEVYTDQDIPGFGVGNLTYESISSTAINVSWVPPAQP 1080
1081 NGLVFFYVSLIILOQTPRHVRPPLVTVYERSIYFDNLEKTYDYLKLTPESTEGFSDVYTAQ 1140
1081 NGLVFFYVSLIILOQTPRHVRPPLVTVYERSIYFDNLEKTYDYLKLTPESTEGFSDVYTAQ 1140
1141 LYKTEEDVPETSPINTFNLSSTSVLLSWDPVPKPGCAIISYDYLQGPENYSPITS 1200
1141 LYKTEEDVPETSPINTFNLSSTSVLLSWDPVPKPGCAIISYDYLQGPENYSPITS 1200
1201 DNVIIILEESPLFLYSFFAAARTRKGLGPSSILFFYTDSEVPLAPQNLTLINCTSDFW 1260
1201 DNVIIILEESPLFLYSFFAAARTRKGLGPSSILFFYTDSEVPLAPQNLTLINCTSDFW 1260
1261 LKMSPSPLPGGIVKVSFKIHEHETDTIYYKNIISGFKTEAKLVGLFPVSTYSIRVSAFTK 1320
1261 LKMSPSPLPGGIVKVSFKIHEHETDTIYYKNIISGFKTEAKLVGLFPVSTYSIRVSAFTK 1320
1321 VGNQGPSNVVKTQESVDPVQVQNCMAWTSQSVLVKWDPPKANGIIITQMTVVERN 1380
1321 VGNQGPSNVVKTQESVDPVQVQNCMAWTSQSVLVKWDPPKANGIIITQMTVVERN 1380
1381 STKVSPQDHMYTFIKLLANTS YVFKVRASTAGDEGDECHVSTLTPETVPSVPTNIAFSD 1440
1381 STKVSPQDHMYTFIKLLANTS YVFKVRASTAGDEGDECHVSTLTPETVPSVPTNIAFSD 1440
1441 VOSTSATLTWIRPDTILGYFQNVKITTLQRAQCKEWESEECVEYQKIQVLYEAHLTEET 1500
1441 VOSTSATLTWIRPDTILGYFQNVKITTLQRAQCKEWESEECVEYQKIQVLYEAHLTEET 1500
1501 VYGLKKFRWYRFQVAASTNAGYGNASNWIISTKTLPGPPDGPENVHVATSPSISISWS 1560
1501 VYGLKKFRWYRFQVAASTNAGYGNASNWIISTKTLPGPPDGPENVHVATSPSISISWS 1560
1561 EBAVITGPTCYLIDVKSVDNDEFNISFIKSNENKTIIEIKOLEIFRYSVVIPTAFTGNIS 1620
1561 EBAVITGPTCYLIDVKSVDNDEFNISFIKSNENKTIIEIKOLEIFRYSVVIPTAFTGNIS 1620
1621 AAYVEGKSAEMIITVTLBSAPKDPNNMTFQKIPDBVTQFQLTFLPPSPNGNIQVYQAL 1680
1621 AAYVEGKSAEMIITVTLBSAPKDPNNMTFQKIPDBVTQFQLTFLPPSPNGNIQVYQAL 1680
1681 VYREDDPTAVQIHNLSIIQKTNFTVIAMLEGLKGGHTYNIISVYVNSAGAGPKVPMRITM 1740
1681 VYREDDPTAVQIHNLSIIQKTNFTVIAMLEGLKGGHTYNIISVYVNSAGAGPKVPMRITM 1740
1741 DIKAPARPKTKPTPIYDATGKLLVSTTTITRMPICYSDDHGPIKQVQLATETGAQHD 1800
1741 DIKAPARPKTKPTPIYDATGKLLVSTTTITRMPICYSDDHGPIKQVQLATETGAQHD 1800

QY 1801 GNVTKWYDAYFNKARPYFTNEGPPNPCTEGKTKFSGNEEIIYIIGADNACMIPEGNEDKIC 1860
Db 1801 GNVTKWYDAYFNKARPYFTNEGPPNPCTEGKTKFSGNEEIIYIIGADNACMIPEGNEDKIC 1860
QY 1861 NGPLPKKKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCILSIILL 1920
Db 1861 NGPLPKKKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCILSIILL 1920
QY 1921 GTAIFAFARIRKQKGGYTSPODAEIIIDTKLKLQDLITVADLEKDERLTRPISSKSP 1980
Db 1921 GTAIFAFARIRKQKGGYTSPODAEIIIDTKLKLQDLITVADLEKDERLTRPISSKSP 1980
QY 1981 QHVEELCTNNLKFQEFSELPKFLQDLSSTDADLPWNAKKNRFPNKPNNNNRVKLI 2040
Db 1981 QHVEELCTNNLKFQEFSELPKFLQDLSSTDADLPWNAKKNRFPNKPNNNNRVKLI 2040
QY 2041 DASVPGSDYINASYISGYLCPNEFIATQGPLGTVGDFWRMVWETRAKTLVMLTQCFEKG 2100
Db 2041 DASVPGSDYINASYISGYLCPNEFIATQGPLGTVGDFWRMVWETRAKTLVMLTQCFEKG 2100
QY 2101 RIRCHQWPEDNKPVTVFQDVIITKLMEDVQIDWTIRDLKIERHGDGMTVROCNFTAWPE 2160
Db 2101 RIRCHQWPEDNKPVTVFQDVIITKLMEDVQIDWTIRDLKIERHGDGMTVROCNFTAWPE 2160
QY 2161 HGVPEASAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220
Db 2161 HGVPEASAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVFTALDHLTOHINDHDFVDI 2220
QY 2221 YGLVAELRSEKMCWONLAQYIPLHOCILDLSNKGSNQPICFVNYSAQKQDMSLDAMEG 2280
Db 2221 YGLVAELRSEKMCWONLAQYIPLHOCILDLSNKGSNQPICFVNYSAQKQDMSLDAMEG 2280
QY 2281 DVELEWEETM 2291
Db 2281 DVELEWEETM 2291

RESULT 3
US-10-087-684-10
; Sequence 10, Application US/10087684
; Publication No. US20040029116A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: MacDougall, John R.
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David J.
; APPLICANT: Grose, William M.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie, J.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Esha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29

; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 10
; LENGTH: 2300
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-684-10

Query Match 98.5%; Score 11807.5; DB 15; Length 2300;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2276; Conservative 5; Mismatches 5; Indels 19; Gaps 9;
QY 1 MDFLIIFLLFIQTSETQVDVSNVVPGRYDITISSITSTYTPVTRIVTPNVPKGPVP 60
Db 1 MDFLIIFLLFIQTSETQVDVSNVVPGRYDITISSITSTYTPVTRIVTPNVPKGPVP 60
QY 61 FLAGERVGSAGILLSNWTPPNNGRIISYIVKYKEVCPMMQVYTVQVRSKPSLEVLN 120
Db 61 FLAGERVGSAGILLSNWTPPNNGRIISYIVKYKEVCPMMQVYTVQVRSKPSLEVLN 120
QY 121 LNPQTTYEIKVAEENSAGIGVDFPDLFQTAES-APGKVNLTVZAYNASAVKLIWYLP 179
Db 121 LNPQTTYEIKVAEENSAGIGVDFPDLFQTAESAPGKVNLTVZAYNASAVKLIWYLP 180
QY 180 QPNKITSFKISVKHARSIGVVKDYSIRVEDILTKLPECN-ENSEFSLWSTASPSPTLG 238
Db 181 QPNKITSFKISVKHARSIGVVKDYSIRVEDILTKLPECNVENSEFLWSTASPSPTLG 240
QY 239 RVTPPSRTHSSSTLTQNEHISWKEPISFVTHLRPPTYLFEVSAATTEAGYIDSTIV 298
Db 241 RVTPPSRTHSSSTLTQNEHISWKEPISFVTHLRPPTYLFEVSAATTEAGYIDSTIV 300
QY 299 RTPESVPEGPONCVTGNITGKSFILWDPPTIVTKGSYRVELYGPSGRILDNSTKOLK 358
Db 301 RTPESVPEGPONCVTGNITGKSFILWDPPTIVTKGSYRVELYGPSGRILDNSTKOLK 360
QY 359 FAPTNLTPPTMYDVYIAAETSAGTGPKSNISVFTPPDVPDGVDFDLQAEVESTQVRITWK 418
Db 361 FAPTNLTPPTMYDVYIAAETSAGTGPKSNISVFTPPDVPDGVDFDLQAEVESTQVRITWK 420
QY 419 KRPQNGIINQVRKVLVPETGIIILENTLLTGNNEYINDPMAPEIINIVPEPMVGLYEGSA 478
Db 421 KRPQNGIINQVRKVLVPETGIIILENTLLTGNNE-INDPMAPEIINIVPEPMVGLYEGSA 479
QY 479 EMSSDLHSLATFIYNSHDPKPNARNABDQTSVVVTTNRQYITDIAAQLSVIRRLVP 538
Db 480 EMSSDLHSLATFIYNSHDPKPNARNABDQTSVVVTTNRQYITDIAAQLSVIRRLVP 539
QY 539 FTEHMSVSAFTIMGEGPPTVLSVTRTQOVPSIKIINYKNISSSILLYWDPPEYPNCK 598
Db 540 FTEHMSVSAFTIMGEGPPTVLSVTRTQOVPSIKIINYKNISSSILLYWDPPEYPNCK 599
QY 599 ITHYTIYAMELDTNRAFOITTTIDNSFLIT--GLKYYTKYKRVAASTHDSSESLSENDI 656
Db 600 ITHYTIYAMELDTNRAFOITTTIDNSFLITIGLKKYTKYKRVAASTHDSSESLSENDI 659
QY 657 FVRTSEDBESSPOQVEVIDVTADRLKWSPEKPNGIIIAYEVLVKNIDTLVKNQVST 716
Db 660 FVRTSEDBESSPOQVEVIDVTADRLKWSPEKPNGIIIAYEVLVKNIDTLVKNQVST 719
QY 717 TDIIILNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI 776
Db 720 TDIIILNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSGEI 779

QY 777 ELSFLPPSSNGIIKKYIYVKRSGNEERTINTSLTONIKVKKYQVHIIIVSASTLK 836
DB 780 ELSFLPPSSNGIIQKYIYVKRSGNEERTINTSLTONI--LKQYQVHIIIVSASTLK 837
QY 837 GEGVRSAPISILTEEDAPSPDQPSVKQLSGVTVKLSWQPLEPENGIIILYTVVYVW-NR 895
DB 838 GEGVRSAPISILTEEDAPSPDQPSVKQLSGVTVKLSWQPLEPENGIIILYTVVYVWR 897
QY 896 SSLKTNVETSLSDLDYVNSYSAVYASTASTFPGDKTGKTSNIIISFQTEGAPSDPPKDV 955
DB 898 SSLKTNVETSLSDLDYVNSYSAVYASTASTFPGDKTGKTSNIIISFQTEG-PSDPPKDV 956
QY 956 YYANLSSSIIILFWTPSPKNGIIQYYSVYRNTSGTFMONTFLHETDNDNMTVSTII 1015
DB 957 YYANLSSSIIILFWTPSPKNGIIQYYSVYRNTSGTFMONTFLHETDNDNMTVSTII 1016
QY 1016 DKLTISYTFWLTASTSVGNKSSDIIEVYTDQDIPFEGFVGNLTYESISSAINVSW 1075
DB 1017 DKLTISYTFWLTASTSVGNKSSDIIEVYTDQDIPFEGFVGNLTYESISSAINVSW 1076
QY 1076 PPAQNGLYVYVSLIIQOTPRHVRPPLVYERSIYFDNLEKTYDILKLTPTSTEGFSD 1135
DB 1077 PPAQNGLYVYVSLIIQOTPRHVRPPLVYERSIYFDNLEKTYDILKLTPTSTEGFSD 1136
QY 1136 TYTAQYIKTEEDVPETSPINTPKNLSSTSVLLSWDPPVKPNGAIIISYDLTLQGNENY 1195
DB 1137 TYTAQYIKTEEDVPETSPINTPKNLSSTSVLLSWDPPVKPNGAIIISYDLTLQGNENY 1196
QY 1196 SPITSDNYIILBLSPTLYSFAAARTKGLGFSIILPYTDESVPPLAPQNLTLINCT 1255
DB 1197 SPITSDNYIILBLSPTLYSFAAARTKGLGFSIILPYTDESVPPLAPQNLTLINCT 1256
QY 1256 SDFVWLKWSPLPGGIVKYVSKIIHEHETDTIYKNIISGFKTEAKLVLEPVSTYSIRV 1315
DB 1257 SDFVWLKWSPLPGGIVKYVSKIIHEHETDTIYKNIISGFKTEAKLVLEPVSTYSIRV 1316
QY 1316 SAFTKVGNGQPSNVKFTTQESVPDVQVQNMCMATSWQSVLVKWDPKKKANGIIITQYVW 1375
DB 1317 SAFTKVGNGQPSNVKFTTQESVPDVQVQNMCMATSWQSVLVKWDPKKKANGIIITQYVW 1376
QY 1376 TVERNSTKVSQDHMTFFIKLLANTSVYFKVRASTAGEGDESTCHVSTLPEVPSVPTN 1435
DB 1377 TVERNSTKVSQDHMTFFIKLLANTSVYFKVRASTAGEGDESTCHVSTLPEVPSVPTN 1436
QY 1436 IAFSDVQSTATLTWIRPDITLGFQNYKITTLQRAQCKEWESECEVYKQIYLYEAH 1495
DB 1437 IAFSDVQSTATLTWIRPDITLGFQNYKITTLQRAQCKEWESECEVYKQIYLYEAH 1496
QY 1496 LTBETVYGLKFRWYRFQVAASNAGYGNASNNISTKTLPGPPDPPENVHVATSPFSI 1555
DB 1497 LTBETVYGLKFRWYRFQVAASNAGYGNASNNISTKTLPGPPDPPENVHVATSPFSI 1556
QY 1556 SISWSEPAVITGTCYLDIVKSVNDNDEFNISPITSKNEENKTIBIKOLEIFTRYSVVITAF 1615
DB 1557 SISWSEPAVITGTCYLDIVKSVNDNDEFNISPITSKNEENKTIBIKOLEIFTRYSVVITAF 1616
QY 1616 TGNISAAVEGKSSAEMIVTLESAPKDPNNNTFQKIDDEVTKFQLTPLPSQPNQNTQ 1675
DB 1617 TGNISAAVEGKSSAEMIVTLESAPKDPNNNTFQKIDDEVTKFQLTPLPSQPNQNTQ 1676
QY 1676 VYQALVREDDPTAVQIHNLISIIQKNTFVIAMLEGLKGGHTNIVSYAVNSAGAGPKVP 1735
DB 1677 VYQALVREDDPTAVQIHNLISIIQKNTFVIAMLEGLKGGHTNIVSYAVNSAGAGPKVP 1736
QY 1736 MRITMDIKAPARKTPTPIYDATGKLLVSTTTITIRMPICYSDDHGPIKNVQVLATET 1795
DB 1737 MRITMDIKAPARKTPTPIYDATGKLLVSTTTITIRMPICYSDDHGPIKNVQVLATET 1796
QY 1796 GAQHDGNVTKWYDAYFNKARPYTNEGFPNPPCTEGKTKFSGNEEIIYIGADNACMIPGN 1855
DB 1797 GAQHDGNVTKWYDAYFNKARPYTNEGFPNPPCTEGKTKFSGNEEIIYIGADNACMIPGN 1856

QY 1856 EDKICNGPLKPKQYLFKFRATNIMQPTDSYSDPVKTLGBGLSRTVEIILSVTLCL 1915
DB 1857 EDKICNGPLKPKQYLFKFRATNIMQPTDSYSDPVKTLGBGLSRTVEIILSVTLCL 1916
QY 1916 SIILIGTALFAPARIRQKQEGTYSPODAEIIDTKLKDQDITVADLELKDBRLTR--- 1972
DB 1917 SIILIGTALFAPARIRQKQEGTYSPODAEIIDTKLKDQDITVADLELKDBRLTRLLS 1976
QY 1973 -----PISKSGFLOHVEELCTNNNLKFOEFSSELPKFLODLSSTADLPWNRKRNFPN 2026
DB 1977 YRKSIIKPIKSGFLOHVEELCTNNNLKFOEFSSELPKFLODLSSTADLPWNRKRNFPN 2036
QY 2027 IKPYNNNRVKLIADASVPGSDIYNASYSIGVLCNPEFATOGPLCTGTGDFRMYWETR 2086
DB 2037 IKPY--NNNRVKLIADASVPGSDIYNASYSIGVLCNPEFATOGPLCTGTGDFRMYWETR 2095
QY 2087 AKTLVMTQCFEGRIRCHQYWPEDNKPTVFGDIVITKLMEDVQIDWTIRDKIERHGD 2146
DB 2096 AKTLVMTQCFEGRIRCHQYWPEDNKPTVFGDIVITKLMEDVQIDWTIRDKIERHGD 2155
QY 2147 CMTVRQCNFTANPEHGVSPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGTGVFIALD 2206
DB 2156 CMTVRQCNFTANPEHGVSPENSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRGTGVFIALD 2215
QY 2207 HUTOHINDHFDVITVGLVAELSRMCMVONLAQYIFLHCQILDLLSNKSGNOPICFVNY 2266
DB 2216 HUTOHINDHFDVITVGLVAELSRMCMVONLAQYIFLHCQILDLLSNKSGNOPICFVNY 2275
QY 2267 SALQKMSLDAMEGDVLEWEETTM 2291
DB 2276 SALQKMSLDAMEGDVLEWEETTM 2300

RESULT 4

US-10-218-779-10

; Sequence 10, Application US/10218779

; Publication No. US20040029222A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit

; APPLICANT: MacDougall, John

; APPLICANT: Millett, Isabelle

; APPLICANT: Ellerman, Karen

; APPLICANT: Stone, David

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William

; APPLICANT: Alsobrook II, John

; APPLICANT: Lepley, Denise

; APPLICANT: Rieger, Daniel

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Spytek, Kimberly

; APPLICANT: Boldog, Perenc

; APPLICANT: Li, Li

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Mishra, Vishnu

; APPLICANT: Patturajan, Meera

; APPLICANT: Shenoy, Suresh

; APPLICANT: Tchernev, Velizar

; APPLICANT: Vernet, Corine

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Malyankar, Uriel

; APPLICANT: Guo, Xiaojia

; APPLICANT: Miller, Charles

; APPLICANT: Gangoli, Eshta

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-214

; CURRENT APPLICATION NUMBER: US/10/218,779

; CURRENT FILING DATE: 2002-08-14

; PRIOR APPLICATION NUMBER: 60/253,834

; PRIOR FILING DATE: 2000-11-29

; PRIOR APPLICATION NUMBER: 60/250,-926

; PRIOR FILING DATE: 2000-11-30

;; PRIORITY APPLICATION NUMBER: 60/264,180
;; PRIORITY FILING DATE: 2001-01-25
;; PRIORITY APPLICATION NUMBER: 60/313,656
;; PRIORITY FILING DATE: 2001-08-20
;; PRIORITY APPLICATION NUMBER: 60/327,456
;; PRIORITY FILING DATE: 2001-10-05
;; NUMBER OF SEQ ID NOS: 216
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 10
;; LENGTH: 2300
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-218-779-10

Query Match 98.5%; Score 11807.5; DB 15; Length 2300;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2276; Conservative 5; Mismatches 5; Indels 19; Gaps 9;
QY 1 MDPLIFLLFLLTGTSETQDVSNVPGTRYDITISISITTYTSPVTRIVTNNVTKPGPV 60
DB 1 MDPLIFLLFLLTGTSETQDVSNVPGTRYDITISISITTYTSPVTRIVTNNVTEGPPV 60
QY 61 FLAGERVGSAGILLSNWTPNPNNGRIISYIVKYKEVCPMMQVYTVQVRSKPSLEVLN 120
DB 61 FLAGERVGSAGILLSNWTPNPNNGRIISYIVKYKEVCPMMQVYTVQVRSKPSLEVLN 120
QY 121 INPGTYIEIKVAENSAGIGVSDPELPQTAES-APGKVNLTVENYASAVKLIWYLR 179
DB 121 INPGTYIEIKVAENSAGIGVSDPELPQTAESAPGKVNLTVENYASAVKLIWYLR 180
QY 180 QNGKITSPKISVGHARSIGVWVDSIRVEDILTGLPCN-ENGSEFLWSTASPTLG 238
DB 181 QNGKITSPKISVGHARSIGVWVDSIRVEDILTGLPCNENVESEFLWSTASPTLG 240
QY 239 RVTPEPRTTHSSSTLTQNEISSVWKEPISFVWVTHLRPYTYLFEVSAATTEAGYIDSTIV 298
DB 241 RVTPEPRTTHSSSTLTQNEISSVWKEPISFVWVTHLRPYTYLFEVSAATTEAGYIDSTIV 300
QY 299 RTPESVPEGPONCVNCTGNITGKSFSLMDPPTIVTGKFSYRVELYGPSRIILDNSTKDLK 358
DB 301 RTPESVPEGPONCVNCTGNITGKSFSLMDPPTIVTGKFSYRVELYGPSRIILDNSTKDLK 360
QY 359 FAFNTLTPPTMDVYIAASTAGTGPKSNI SVFTPDVPGVDFDLQAEVSTQVRIWTK 418
DB 361 FAFNTLTPPTMDVYIAASTAGTGPKSNI SVFTPDVPGVDFDLQAEVSTQVRIWTK 420
QY 419 KPRQNGIINQYRVKLVLPETGII LENTLLTGNEYINDPMAPEI NVIPEVMVGLYEGSA 478
DB 421 KPRQNGIINQYRVKLVLPETGII LENTLLTGNE-INDPMAPEI NVIPEVMVGLYEGSA 479
QY 479 EMSSDLHSLATFIYNSHPDKNFPARNRABDQTSVVTTRNQVITDIAABQLSVYRRLVP 538
DB 480 EMSSDLHSLATFIYNSHPDKNFPARNRABDQTSVVTTRNQVITDIAABQLSVYRRLVP 539
QY 539 FTEHMSVSAFTIMGEGPPTLVSRTRQOVPSIKIINVKNISSSILLYMDPPYVNGK 598
DB 540 FTEHMSVSAFTIMGEGPPTLVSRTRQOVPSIKIINVKNISSSILLYMDPPYVNGK 599
QY 599 ITHYTIYAMELDTNRAFOITTDNSPLIT--GLKKYTKYKMRVAASTHDSGSSLEENDI 656
DB 600 ITHYTIYAMELDTNRAFOITTDNSPLITGIGLKKYTKYKMRVAASTHDSGSSLEENDI 659
QY 657 FVRTSEDESPQDVEVDVTADEIRLWSPPEKNGII IAYEVLVYKNIIDLYMKNTST 716
DB 660 FVRTSEDESPQDVEVDVTADEIRLWSPPEKNGII IAYEVLVYKNIIDLYMKNTST 719
QY 717 TDIILNLNPHLTYNLSVRSYTRFGHGNQVSSILSVRTSETVPSAPENITVKNISSGEI 776
DB 720 TDIILNLNPHLTYNLSVRSYTRFGHGNQVSSILSVRTSETVPSAPENITVKNISSGEI 779
QY 777 ELSFLPPSPNGI IKKYTIYLRKSNGBERTINTTSLTQNIKVLYKQYIIEVSASTLK 836
DB 780 ELSFLPPSPNGI IKKYTIYLRKSNGBERTINTTSLTQNI--LKKYTYIIEVSASTLK 837

QY 837 GEGVRSAPISILTEEDAPDPSPPQDPSVKQLSGVTVKLSWQPPLEPNGIILYTVVW-NR 895
DB 838 GEGVRSAPISILTEEDAPDPSPPQDPSVKQLSGVTVKLSWQPPLEPNGIILYTVVWNR 897
QY 896 SSLKTIINVTETSLSDLDYNYEYSAVYTASTRFGDGKTSNIIISFQTPGAPSPKDV 955
DB 898 SSLKTIINVTETSLSDLDYNYEYSAVYTASTRFGDGKTSNIIISFQTPG-PSOPKDV 956
QY 956 YYANLSSSSIIILFWTPPSKPGCI IQYYSVYRNTSGTFMONTLHELINDFNMTVSTII 1015
DB 957 YYANLSSSSIIILFWTPPSKPGCI IQYYSVYRNTSGTFMONTLHELINDFNMTVSTII 1016
QY 1016 DKLTIFSYTYFWLTASTSVGNKSSDIIEVYTDODIPEGFVGNITYESISSTAINVSV 1075
DB 1017 DKLTIFSYTYFWLTASTSVGNKSSDIIEVYTDQDVPEGFVGNITYESISSTAINVSV 1076
QY 1076 PPAQNGLVFVYVSHILOQTPRHVRPPLVYERSIYFDNLEKYTDIILKITPSTKGFSD 1135
DB 1077 PPAQNGLVFVYVSHILOQTPRHVRPPLVYERSIYFDNLEKYTDIILKITPSTKGFSD 1136
QY 1136 TYTAQLYIKTEEDVPETSPIINTFKNLSSTSVLLSWDPPVKPENGAIISYDLTLQPNENY 1195
DB 1137 TYTAQLYIKTEEDVPETSPIINTFKNLSSTSVLLSWDPPVKPENGAIISYDLTLQPNENY 1196
QY 1196 SPITSDNYIIIEELSPPFTLYSFFFAARTKGI GPSSILFFYTDSEVPPLAPPONLTLINCT 1255
DB 1197 SPITSDNYIIIEELSPPFTLYSFFFAARTKGI GPSSILFFYTDSEVPPLAPPONLTLINCT 1256
QY 1256 SDFVWLKWSPLPGGIVKVYSFKIHEHETDTIYKNISGFKTEAKLVGLEPVSTYSIRV 1315
DB 1257 SDFVWLKWSPLPGGIVKVYSFKIHEHETDTIYKNISGFKTEAKLVGLEPVSTYSIRV 1316
QY 1316 SAFTKVGNGNQPNNVKTTOESVDDVQNNQCMATSWQSVLVKWDPPKANGIITQYV 1375
DB 1317 SAFTKVGNGNQPNNVKTTOESVDDVQNNQCMATSWQSVLVKWDPPKANGIITQYV 1376
QY 1376 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGEDESTCHVSTLPSTVSPVPTN 1435
DB 1377 TVERNSTKVSPODHMYTFIKLLANTSYVFKVRASTASAGEDESTCHVSTLPSTVSPVPTN 1436
QY 1436 IAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESECEVEYQKIQLYBAH 1495
DB 1437 IAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQCKEWESECEVEYQKIQLYBAH 1496
QY 1496 LTEETVYGLKPRWTRPOVAASTNAGYGNASWISTKTILPGDPPDPNNVHVVAISPSFI 1555
DB 1497 LTEETVYGLKPRWTRPOVAASTNAGYGNASWISTKTILPGDPPDPNNVHVVAISPSFI 1556
QY 1556 SISWSEPAVITGPTCYLIDVKSVDNDEFNIFIKSNEENKTIIEIKDLEIFTRYSVVITAF 1615
DB 1557 SISWSEPAVITGPTCYLIDVKSVDNDEFNIFIKSNEENKTIIEIKDLEIFTRYSVVITAF 1616
QY 1616 TGNISAAVEGKSSAEMIVTTLESAPDPPNNMTFKIPDEVTQFQLTFLPPSQPNIGIQ 1675
DB 1617 TGNISAAVEGKSSAEMIVTTLESAPDPPNNMTFKIPDEVTQFQLTFLPPSQPNIGIQ 1676
QY 1676 VYQALVYREDPTAVQIHNLSIIQKNTFTVIAMLEGLKGHTYNTISVAVNSAGAPKVP 1735
DB 1677 VYQALVYREDPTAVQIHNLSIIQKNTFTVIAMLEGLKGHTYNTISVAVNSAGAPKVP 1736
QY 1736 MRLTMDIKAPARKTKPTPIYDATGKLVSTTTITRMPICVYSDHGPKNQVILATFT 1795
DB 1737 MRLTMDIKAPARKTKPTPIYDATGKLVSTTTITRMPICVYSDHGPKNQVILATFT 1796
QY 1796 GAQHDGNTYKWDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEBIIYIIGADNACMIPGN 1855
DB 1797 GAQHDGNTYKWDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEBIIYIIGADNACMIPGN 1856
QY 1856 EDKICNGPLKPKQVLFKFRATNIMGQTSDDSDYSPVKTILGEGLSERTVEIILSVTLCIL 1915
DB 1857 EDKICNGPLKPKQVLFKFRATNIMGQTSDDSDYSPVKTILGEGLSERTVEIILSVTLCIL 1916

```
QY 1916 SIILLGTAIPAFARIKOKEGTYSPODAEIIDTKLKDOLITVADLEIKDRLTR--- 1972
Db 1917 SIILLGTAIPAFARIKOKEGTYSPODAEIIDTKLKDOLITVADLEIKDRLTRLS 1976
QY 1973 -----PISKSGFLOHVEBELCTNNNLKFOEBSFSELPKFLQDLSSSTDADLPWRAKORFPN 2026
Db 1977 YRKSIRKPIKSGSGFLOHVEBELCTNNNLKFOEBSFSELPKFLQDLSSSTDADLPWRAKORFPN 2036
QY 2027 IKPYNNNRVKLIADASVGSYDYNASYSIGVLCNPEFTATQGPLGTGDFWVMWETR 2086
Db 2037 IKPY-NNNRVKLIADASVGSYDYNASYSIGVLCNPEFTATQGPLGTGDFWVMWETR 2095
QY 2087 AKTLVMLTQCFKGRIRCHOYMPEDNKPTVFGDIVITKLMEDVQIDWTIRDLKIERHGD 2146
Db 2096 AKTLVMLTQCFKGRIRCHOYMPEDNKPTVFGDIVITKLMEDVQIDWTIRDLKIERHGD 2155
QY 2147 CMTVROCNTPAPEHGVSPENSAPLIHFVKLVASRAHDTTPMLVHCSAGVGRGTGVFIAD 2206
Db 2156 CMTVROCNTPAPEHGVSPENSAPLIHFVKLVASRAHDTTPMLVHCSAGVGRGTGVFIAD 2215
QY 2207 HLTOHNDHDFVDIYGLVAELRSERCMVONLAQYIFLHCILDLNKGNSNQPICFVNY 2266
Db 2216 HLTOHNDHDFVDIYGLVAELRSERCMVONLAQYIFLHCILDLNKGNSNQPICFVNY 2275
QY 2267 SALQKMDSLDAMEGDVELEWEETM 2291
Db 2276 SALQKMDSLDAMEGDVELEWEETM 2300
```

RESULT 5

```
US-10-466-759-2
; Sequence 2, Application US/10466759
; Publication No. US20040081983A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: LEE, Ernestine A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: ISON, Craig H.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: JACKSON, Jennifer L.
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: TRAN, Bao
; APPLICANT: DING, Li
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: WARREN, Bridget A.
; TITLE OF INVENTION: KINASES AND PHOSPHATASES
; FILE REFERENCE: PI-0344 USN
; CURRENT APPLICATION NUMBER: US/10/466,759
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: PCT/US02/01369
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: US 60/263,083
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/271,205
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/271,117
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/276,859
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/278,504
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/278,522
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/280,510
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/280,266
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
```

```
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 2299
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7480588CD1
US-10-466-759-2

Query Match          98.4%; Score 11797; DB 15; Length 2299;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2272; Conservative 6; Mismatches 5; Indels 24; Gaps 8;

QY 1 MDFLIIFLLPFTGTSETQVDVSNVVPGRYDITISISITTYTSPVTRIIVTPNVTKEGPPV 60
Db 1 MDFLIIFLLPFTGTSETQVDVSNVVPGRYDITISISITTYTSPVTRIIVTPNVTKEGPPV 55
QY 61 FLAGERSVGSAGILLSNWTPPNNGRIISYIVKYKEVCPMMQVYTVQVRSKPSDLEVLN 120
Db 56 FLAGERSVGSAGILLSNWTPPNNGRIISYIVKYKEVCPMMQVYTVQVRSKPSDLEVLN 115
QY 121 LNPGETTYEIKVAEENSAGIGVFSDFLFTQTAESAAGKVVNLTVAEYNASAVKLITWLP 180
Db 116 LNPGETTYEIKVAEENSAGIGVFSDFLFTQTAESAAGKVVNLTVAEYNASAVKLITWLP 175
QY 181 PNGKITSPKISVKHARSGIWKDVIRVEDILTGLPECNENSESFLWSTASPSPTLGRV 240
Db 176 PNGKITSPKISVKHARSGIWKDVIRVEDILTGLPECNENSESFLWSTASPSPTLGRV 235
QY 241 TPPSRTHSSSTLTONEISSWKEPISFVWTHLRPYTTLPEVSAATTEAGYIDSTIVRT 300
Db 236 TPPSRTHSSSTLTONEISSWKEPISFVWTHLRPYTTLPEVSAATTEAGYIDSTIVRT 295
QY 301 PESVPEGPQNCVTCNITGKSPSILLWDPPTIVTGKFSYRVELYGSPS--GRILDNSTKDLK 358
Db 296 PESVPEGPQNCVTCNITGKSPSILLWDPPTIVTGKFSYRVELYGSPS--GRILDNSTKDLK 355
QY 359 PAFNTLTPPTMYDVYIAAETSAGTGPKSGNISVFTPPDVGAVDFDLQAEVSTQVRAITWK 418
Db 356 PAFNTLTPPTMYDVYIAAETSAGTGPKSGNISVFTPPDVGAVDFDLQAEVSTQVRAITWK 415
QY 419 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNNEYINDMPAIEVINIYVPMVGLYEGSA 478
Db 416 KRPQNGIINQYRVKVLVPETGIIILENTLLTGNNE-INDMPAIEVINIYVPMVGLYEGSA 474
QY 479 EMSDDLHSLATPIYNSHPDKNPPARNRAEDQTSPPVVTTRNQYITDIAAQSLSVIRRLVP 538
Db 475 EMSDDLHSLATPIYNSHPDKNPPARNRAEDQTSPPVVTTRNQYITDIAAQSLSVIRRLVP 534
QY 539 FTEHMISSAFTIMGEGPPTVLSVTRQVPSIKIINYKNISSSILLYWDPPEYPNGK 598
Db 535 FTEHMISSAFTIMGEGPPTVLSVTRQVPSIKIINYKNISSSILLYWDPPEYPNGK 594
QY 599 ITHYTIYAMELDTNRAFOITTDNSFLITG-----LKYTKYKMRVAASTHDSSELSSEN 654
Db 595 ITHYTIYAMELDTNRAFOITTDNSFLITGGRKWLKKTYYKMRVAASTHDSSELSSEN 654
QY 655 DIFVTSSEDEPSSQDVEIDVTADEIRLKWSPKPNGLIIAYEVLKNDITLTKMNT 714
Db 655 DIFVTSSEDEPSSQDVEIDVTADEIRLKWSPKPNGLIIAYEVLKNDITLTKMNT 714
QY 715 STTDIILNLRPHLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSPAENITYKNISSG 774
Db 715 STTDIILNLRPHLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSPAENITYKNISSG 774
QY 775 EIELSFLPPSSPNGIIKKYTIYVKRSGNEERTINTTSLTONIKVLKKTQYIIEVSAST 834
Db 775 EIELSFLPPSSPNGIIKKYTIYVKRSGNEERTINTTSLTONIKVLKKTQYIIEVSAST 834
QY 835 LKGEGRVRSAPISILTEEDAPDPPDQFSVKQLSGVTVKLSWQPLEPNCIIILYTYVWV- 893
Db 835 LKGEGRVRSAPISILTEEDAPDPPDQFSVKQLSGVTVKLSWQPLEPNCIIILYTYVWV- 894
```

```
QY 894 NRSSLLKTIIVTETSLDLDYVVEYSAYVTASTRFGDGKTSNIIISFOTPEGAPSDPK 953
Db 895 NRSSLLKTIIVTETSLDLDYVVEYSAYVTASTRFGDGKTSNIIISFOTPEG-PSDPK 953
QY 954 DVYANLSSSSIIILFWTPSPKNGIIQYYSVYRNTSGTFMQNFTLHETNDPDMVTST 1013
Db 954 DVYANLSSSSIIILFWTPSPKNGIIQYYSVYRNTSGTFMQNFTLHETNDPDMVTST 1013
QY 1014 IIDKLTFISYTFWLTASTSVGNKSSDIIIEVYTDQDIPGPFVGNLTYESISSSTAINVS 1073
Db 1014 IIDKLTFISYTFWLTASTSVGNKSSDIIIEVYTDQDIPGPFVGNLTYESISSSTAINVS 1073
QY 1074 WPPAPQNGLVFYVSLIIQQTPRHVRPPLVYVESIYFDNLEKYDYILKLTPTSEKGF 1133
Db 1074 WPPAPQNGLVFYVSLIIQQTPRHVRPPLVYVESIYFDNLEKYDYILKLTPTSEKGF 1133
QY 1134 SDTYTAQLYIKTEEDVPETSPIINTPKNLSSVLLSWDPVPKPGAIISYDLTLQGPNE 1193
Db 1134 SDTYTAQLYIKTEEDVPETSPIINTPKNLSSVLLSWDPVPKPGAIISYDLTLQGPNE 1193
QY 1194 NYSFITSNDYIIIEELSPFTLSYFFAAARTKGLGSSILFFYTBESVPLAPPQNLTLIN 1253
Db 1194 NYSFITSNDYIIIEELSPFTLSYFFAAARTKGLGSSILFFYTBESVPLAPPQNLTLIN 1253
QY 1254 CTSDFVWLKWSPLPGGIVKYVSPKIHETDITIYKNI SGPKTEAKLVGLEPVSTYSI 1313
Db 1254 CTSDFVWLKWSPLPGGIVKYVSPKIHETDITIYKNI SGPKTEAKLVGLEPVSTYSI 1313
QY 1314 RVSFAFTKVGNGQFNVVFTTQESVDPVQNNQCMATSWQSVLVKWDPPKXANGIIQY 1373
Db 1314 RVSFAFTKVGNGQFNVVFTTQESVDPVQNNQCMATSWQSVLVKWDPPKXANGIIQY 1373
QY 1374 MVTVERNSTKVPQDMYTFIKLLANTS YVFKVRASTSGEDSSTCHVSTLPTVPVSP 1433
Db 1374 MVTVERNSTKVPQDMYTFIKLLANTS YVFKVRASTSGEDSSTCHVSTLPTVPVSP 1433
QY 1434 TNLAFSDVOSTATLTIWIPDITLGYFQNYKITQLRACKKEWSEECVEYKQIYLYE 1493
Db 1434 TNLAFSDVOSTATLTIWIPDITLGYFQNYKITQLRACKKEWSEECVEYKQIYLYE 1493
QY 1494 AHLTEBTVGLKFRWYRQVAASTNAGYCNASNISTKTLPQPPGPPENHVWATSPF 1553
Db 1494 AHLTEBTVGLKFRWYRQVAASTNAGYCNASNISTKTLPQPPGPPENHVWATSPF 1553
QY 1554 SISISWSEPAVITGPTCYLIDVKVDNDFNISFKISNENKTIETIKOLEIFTRYSVWIT 1613
Db 1554 SISISWSEPAVITGPTCYLIDVKVDNDFNISFKISNENKTIETIKOLEIFTRYSVWIT 1613
QY 1614 APTGNISAAVVEGKSAEMI VTTLESAPKDPNNMTFQKI PDEVTKFQLTFLPPSPNGN 1673
Db 1614 APTGNISAAVVEGKSAEMI VTTLESAPKDPNNMTFQKI PDEVTKFQLTFLPPSPNGN 1673
QY 1674 IOVYQALVREDDPTAVQIHNLSIIQKNTFTVIALEGLKGHTYNSIYVAVNSAGAPK 1733
Db 1674 IOVYQALVREDDPTAVQIHNLSIIQKNTFTVIALEGLKGHTYNSIYVAVNSAGAPK 1733
QY 1734 VPMRITMDIKAPARPKTPTPIYDATGKLLVSTTITIRMPICYSDDHGPIKNVOVLAT 1793
Db 1734 VPMRITMDIKAPARPKTPTPIYDATGKLLVSTTITIRMPICYSDDHGPIKNVOVLAT 1793
QY 1794 ETGAQHDGNTWYDAYFNKARPYFTNEGFPNPPCTEGTKYSGNEEIIYIIGADNACMIP 1853
Db 1794 ETGAQHDGNTWYDAYFNKARPYFTNEGFPNPPCTEGTKYSGNEEIIYIIGADNACMIP 1853
QY 1854 GNEDKICNGPLKPKQYLFKFRATNIMGQFTDSYDSDPVKTLGEGLSERTVEIILSVTLIC 1913
Db 1854 GNEDKICNGPLKPKQYLFKFRATNIMGQFTDSYDSDPVKTLGEGLSERTVEIILSVTLIC 1913
QY 1914 ILSIILLGTAFAPARIRQKQEGGYSPQDAEIIDTKLKLQDLITVADLELKDRLTR- 1972
Db 1914 ILSIILLGTAFAPARIRQKQEGGYSPQDAEIIDTKLKLQDLITVADLELKDRLTR- 1973

QY 1973 -----PISKSKFLQHVHEELCTNNNLKFQBEFSELPKFLQDLSSTADLPWNAKQNF 2024
Db 1974 LSYRSKSIKPIKSKSKFLQHVHEELCTNNNLKFQBEFSELPKFLQDLSSTADLPWNAKQNF 2033
QY 2025 PNKPYNNNRVVKLTADASVPGSDYINASYISGYLCNPEFIATQGPLPGTVGD FWRMVWE 2084
Db 2034 PNKEY-NNNRVKLTADASVPGSDYINASYISGYLCNPEFIATQGPLPGTVGD FWRMVWE 2092
QY 2085 TRAKTLVMLTQCFEKGRI RCHQYWPEDNKPVTVPFGDIVITKLMEVDQIDWTIRDLKIERH 2144
Db 2093 TRAKTLVMLTQCFEKGRI RCHQYWPEDNKPVTVPFGDIVITKLMEVDQIDWTIRDLKIERH 2152
QY 2145 GDCMTRVRCNFTAWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVFIA 2204
Db 2153 GDCMTRVRCNFTAWPEHGVPENSAPLIHFVKLVASRAHDTTPMIVHCSAGVGRGVFIA 2212
QY 2205 LDHLTQHINDHDFVDIYGLVAELRSERCMQVNLQAYIFLHQICILDLSNKGSNQPICFV 2264
Db 2213 LDHLTQHINDHDFVDIYGLVAELRSERCMQVNLQAYIFLHQICILDLSNKGSNQPICFV 2272
QY 2265 NYSALQKQSDSLDAMEGDVELEWEETTM 2291
Db 2273 NYSALQKQSDSLDAMEGDVELEWEETTM 2299

RESULT 6
US-10-087-684-6
; Sequence 6, Application US/10087684
; Publication NO. US20040029116A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: MacDougall, John R.
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David J.
; APPLICANT: Grosse, William M.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgesse, Catherine E.
; APPLICANT: Casman, Stacie, J.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Li, Li
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Esha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 6
```

; LENGTH: 2281

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-087-684-6

Query Match 93.4%; Score 11191; DB 15; Length 2281;
Best Local Similarity 94.5%; Pred. No. 0;
Matches 217; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

QY 1 MDLFIPLLLFIGTSETQDVSNVPGTRVYDITISSITTTTSPVTRIVPNVTKGPPV 60
DB 1 MDLFIPLLLFIGTSETQDVSNVPGTRVYDITISSITTTTSPVTRIGASN--EPGPPV 58
QY 61 FLAGERVGSAGILLSNWTPNPNNGRIISYIVKVEVCPNMQVYTVQVRSXPDSLEVLITN 120
DB 59 FLAGERVGSAGILLSNWTPNPNNGRIISYIVKVEVCPNMQVYTVQVRSXPDSLEVLITN 118
QY 121 LNPCTTVEIKVAEASAGIVGDFDPFLQFAESAPGKVNLTVAEYNASAVKLIWLPQ 180
DB 119 LNPCTTVEIKVAEASAGIVGDFDPFLQFAESAPGKVNDFTGEAVPFSS--KLMWTSAT 177
QY 181 PNGKITSFKISVKHARSIGVVKDVSIRVEDILTGLP-ECNENSESFLMSTASPSPTLGR 239
DB 178 KK-KITSFKISVKHRSIGVIVKEVSIRVECILSASLPHCNENSESFLMSTASPSPTLGR 236
QY 240 VTPPSRTHSSSTLTONEISSVWKEPISPVVTTHLRPVYTVLFFVSAATTEAGYIDSTIVR 299
DB 237 VTPPSRTHSSSTLTONEISSV--KEPISFVVVTHLRPVYTVLFFVSAATTEAGYIDSTIVR 295
QY 300 TPESVPEGPONCVTGNITCKSILWDPPTIVTGKFSYRVELYGPS-GRILDNSTKDLK 358
DB 296 TPESVPEGPONCVTGNITCKSILWDPPTIVTGKFSYRVELYGPSAGRILDNSTKDLK 355
QY 359 FATNLTPFTMYDVYIAAETSAGTGPKSNISVTPDPVCAVFDLQLAEVESQVRIITWK 418
DB 356 FATNLTPFTMYDVYIAAETSAGTGPKSNISVTPDPVCAVFDLQLAEVESQVRIITWK 415
QY 419 KRPQNGIINOYRKVKLVPTGTIILENTLLTGNNVINDPMAPEIIVNIEPVMVGLYEGSA 478
DB 416 KRPQNGIINOYRKVKLVPTGTIILENTLLTGNN--INDMAPEIIVNIEPVMVGLYEGSA 474
QY 479 EMSSDLHSLATFYNSHPDKNFPARNRAEDQTSFVVTTRNOYITDIAEQLSVIRRLVP 538
DB 475 EMSSDLHSLATFYNSHPDKNFPARNRAEDQTSFVVTTRNOYITDIAEQLSVIRRLR 534
QY 539 FTEHMLSVSAFTIMGEGPPTVLSVRTRQOVPSIKIINLK--NISSSILLYWDPPPEYN 596
DB 535 FNAETGFSRYITIMSS-----ASRDNLTSPGLSAQNFRVTHVITTEVFLHWDPPD--P 586
QY 597 GKITHYTIYAMELDT-NRAFOITIDN-SFLITGLKKYTKYKRVAAASTHDGESSISEEN 654
DB 587 VPFHHVLTILDVENQSKSILATLNSLSLVGLKKYTKYKRVAAASTHVGESSISEEN 646
QY 655 DIFVRSDEBPSSQDVEIDVTADIRLKWSPPEKPGNIIIAEVLVKNIDTLVKNKT 714
DB 647 DIFVRSDEBPSSQDVEIDVTADIRLKWSPPEKPGNIIIAEVLVKNIDTLVKNKT 706
QY 715 STTDIILRNLRPHTLNYSVRSYTRCHGNQVSSLSVRSVTPDASAPENITYKNISG 774
DB 707 STTDIILRNLRPHTLNYSVRSYTRCHGNQVSSLSVRSVTPDASAPENITYKNISG 766
QY 775 EIELSELFPSSNGIITKYYIYLRSGNBERINTTSLTONIKVLKKTQYIIEVSAST 834
DB 767 EIELSELFPSSNGIITKYYIYLRSGNBERINTTSLTONIKVLKKTQYIIEVSAST 826
QY 835 LKGEGRVSRAPISILTEEDAPSPQDFSVKQLSGVTVKLSQWPLPENGIIYYTYVWN 894
DB 827 LKGEGRVSRAPISILTEEDAPSPQDFSVKQLSGVTVKLSQWPLPENGIIYYTYVWN 885
QY 895 RSSLKKTINVTETSLDLDYNYEYSAVYTASTRFGDKTGSNIISFQTPPEGAPSPPKD 954
DB 886 RSSLKKTINVTETSLDLDYNYEYSAVYTASTRFGDKTGSNIISFQTPPEG-PSDPPPKD 944

QY 955 VYIANLSSSSIIILFWTPPSKPNGIIQYYSVYVYRNTSGTFMONTLHELTFNDPDMTVSTI 1014
DB 945 VYIANLSSSSIIILFWTPPSKPNGIIQYYSVYVYRNTSGTFMONTLHELTFNDPDMTVSTI 1004
QY 1015 IDKLTIFSYTFLWTASTSVGNKSSDIEVYTDQDIPGFGVGNLTYSISSTAINVSW 1074
DB 1005 IDKLTIFSYTFLWTASTSVGNKSSDIEVYTDQDIPGFGVGNLTYSISSTAINVSW 1064
QY 1075 VPPAQNGLVFYVYSIILQOTPRHVRPPLVTVYERSYIFDNLEKYTYIILKITESTKGF 1134
DB 1065 VPPAQNGLVFYVYSIILQOTPRHVRPPLVTVYERSYIFDNLEKYTYIILKITESTKGF 1124
QY 1135 DTYTAQLYIKTEEDVPETSPINTFNKLSSTSVLLSWDPVVKNGAIISYDLTLQGNEN 1194
DB 1125 DTYTAQLYIKTEEDIPETSPINTFNKLSSTSVLLSWDPVVKNGAIISYDLTLQGNEN 1184
QY 1195 YSFITSDNVIILEELSPFTLYSFFFAAARTRKGLGPSIILFFYTDSEVPLAPPQNLTINC 1254
DB 1185 YSFITSDNVIILEELSPFTLYSFFFAAARTRKGLGPSIILFFYTDSEVPLAPPQNLTINC 1244
QY 1255 TSDFVWLKWSPSPLPGGIVKVSFKIHEHETDTIYKNIISGPKTEAKLVGLEPVSTYSIR 1314
DB 1245 TSDFVWLKWSPSPLPGGIVKVSFKIHEHETDTIYKNIISGPKTEAKLVGLEPVSTYSIR 1304
QY 1315 VSAFTKVGNGNQSNNVVKTTQESVDPVQVQNCMAATSWQSVLVKWDPPKANGIITQYM 1374
DB 1305 VSAFTKVGNGNQSNNVVKTTQESVDPVQVQNCMAATSWQSVLVKWDPPKANGIITQYM 1364
QY 1375 VTVERNSTKVSQDQHMVTEIKLLANTSYVFKVRASTASAGDESTCHVSTLPETVSVPT 1434
DB 1365 VTVERNSTKVSQDQHMVTEIKLLANTSYVFKVRASTASAGDESTCHVSTLPETVSVPT 1424
QY 1435 NIAFSDVQSTATLWIRPDTILGYFQNTKIITQLRAQCKWESECEVEYQKIQLYLEA 1494
DB 1425 NIAFSDVQSTATLWIRPDTILGYFQNTKIITQLRAQCKWESECEVEYQKIQLYLEA 1484
QY 1495 HLTEETVYGLKFRWTRFOVAASTNAGYNASNISTKTLPGPDGPPENHVAVTSPPS 1554
DB 1485 HLTEETVYGLKFRWTRFOVAASTNAGYNASNISTKTLPGPDGPPENHVAVTSPPS 1544
QY 1555 ISISSEBPAVITGPTCYLIDVKSVDNDEFNIFSPIKSEENKTIIRIKDLIFTRYSVVITA 1614
DB 1545 ISISSEBPAVITGPTCYLIDVKSVDNDEFNIFSPIKSEENKTIIRIKDLIFTRYSVVITA 1604
QY 1615 FTGNTSAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPPSQNGNI 1674
DB 1605 FTGNTSAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPPSQNGNI 1664
QY 1675 QYVQALVYREDDPTAVQIHNLISIIQKNTFTVIAMLEGLKGGHTYINSVYAVNSAGAPKV 1734
DB 1665 QYVQALVYREDDPTAVQIHNLISIIQKNTFTVIAMLEGLKGGHTYINSVYAVNSAGAPKV 1724
QY 1735 PMRITMDIKAPARPKTPTPIYDATGKLLVTSITTIIRMPICIYSDDHGPIKVVQVLAIE 1794
DB 1725 PMRITMDIKAPARPKTPTPIYDATGKLLVTSITTIIRMPICIYSDDHGPIKVVQVLAIE 1784
QY 1795 TGAQHDGNTYKWDAYFNKARPYFTNEGPNPCTEGKTKFSGNEBIYIIGADNACWIPG 1854
DB 1785 TGAQHDGNTYKWDAYFNKARPYFTNEGPNPCTEGKTKFSGNEBIYIIGADNACWIPG 1844
QY 1855 NEDKICNGPLKPKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCI 1914
DB 1845 NEDKICNGPLKPKQYLFKFRATNIMGQFTDSYSDPVKTLGEGLSERTVEIILSVTLCI 1904
QY 1915 LSIILLLGTAIFAFARIQKQEGGTYSPDABIIIDTKLKDQITVADLELKDRLTRPI 1974
DB 1905 LSIILLLGTAIFAFARIQKQEGGTYSPDABIIIDTKLKDQITVADLELKDRLTRPI 1964
QY 1975 SKKSFLOHVEELCTNNLKFQEFSELPKFLODLSSTDADLPWNRANKFPNPKYNNNN 2034
DB 1965 SKKSFLOHVEELCTNNLKFQEFSELPKFLODLSSTDADLPWNRANKFPNPKYNNNN 2023
QY 2035 RVKLIADASVPGSDYINASYISGYLCPNBEFIATQGPLCTGVGDFWRMNVWETRAKTLVMLT 2094

Db 2024 RVKLIADASVPCSDYINASYISGYLCPNEFIATQGLPGTGVDFWVWVETRAKTLVMLT 2083
QY 2095 QCFEKGRIRCHQWPDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRQCN 2154
Db 2084 QCFEKGRIRCHQWPDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDGMTVRQCN 2143
QY 2155 FTAMPBHGVPENSAPLIHFVKLVASRAHDTPPMIVHCSAGVGRGTGFALDHLTOHIND 2214
Db 2144 FTAMPBHGVPENSAPLIHFVKLVASRAHDTPPMIVHCSAGVGRGTGFALDHLTOHIND 2203
QY 2215 HDFVDIYGLVAELSRMCMQVQLAQYIFLHQCILDLLSNKSGNQICFVNTSALQXKMD 2274
Db 2204 HDFVDIYGLVAELSRMCMQVQLAQYIFLHQCILDLLSNKSGNQICFVNTSALQXKMD 2263
QY 2275 LDAME-GDVELEWEETM 2291
Db 2264 LDAMEGGDVELEWEETM 2281

RESULT 7

US-10-218-779-6

/ Sequence 6, Application US/10218779
/ Publication No. US20040029222A1

GENERAL INFORMATION:

/ APPLICANT: Edinger, Shlomit
/ APPLICANT: MacDougall, John
/ APPLICANT: Millet, Isabelle
/ APPLICANT: Ellerman, Karen
/ APPLICANT: Stone, David
/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Grose, William
/ APPLICANT: Alsbrook II, John
/ APPLICANT: Lepley, Denise
/ APPLICANT: Rieger, Daniel
/ APPLICANT: Burgess, Catherine
/ APPLICANT: Casman, Stacie
/ APPLICANT: Spytek, Kimberly
/ APPLICANT: Boldog, Ferenc
/ APPLICANT: Li, Li
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Mishra, Vishnu
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Shenoy, Suresh
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Tchernev, Velizar
/ APPLICANT: Vernet, Corine
/ APPLICANT: Zerhusen, Bryan
/ APPLICANT: Malyankar, Uriel
/ APPLICANT: Guo, Xiaojia
/ APPLICANT: Miller, Charles
/ APPLICANT: Gangolli, Esha

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

/ FILE REFERENCE: 21402-214

/ CURRENT APPLICATION NUMBER: US/10/218,779

/ CURRENT FILING DATE: 2002-08-14

/ PRIOR APPLICATION NUMBER: 60/253,834

/ PRIOR FILING DATE: 2000-11-29

/ PRIOR APPLICATION NUMBER: 60/250,-926

/ PRIOR FILING DATE: 2000-11-30

/ PRIOR APPLICATION NUMBER: 60/264,180

/ PRIOR FILING DATE: 2001-01-25

/ PRIOR APPLICATION NUMBER: 60/313,656

/ PRIOR FILING DATE: 2001-08-20

/ PRIOR APPLICATION NUMBER: 60/327,456

/ PRIOR FILING DATE: 2001-10-05

/ NUMBER OF SEQ ID NOS: 216

/ SOFTWARE: Patentin Ver. 2.1

/ SEQ ID NO 6

/ LENGTH: 2281

/ TYPE: PRT

/ ORGANISM: Homo sapiens

US-10-218-779-6

Query Match 93.4%; Score 11191; DB 15; Length 2281;
Best Local Similarity 94.5%; Pred. No. 0;
Matches 2171; Conservative 40; Mismatches 63; Indels 24; Gaps 16;

QY 1 MDPLIIFLLFLIGTSETQVDVSNVVPGRYDITISSISTYTSPTVTRIVTPNVTKEGPPV 60
Db 1 MDPLIIFLLFLIGTSETQVDVSNVVPGRYDITISSISTYTSPTVTRIGASN--EKGPPV 58

QY 61 FLAGERVSGAGILLSWNTPPNPNGRISIVIVKYKEVCPMMQTVYTVQVRSKPSLEVLITN 120
Db 59 FLAGERVSGAGILLSWNTPPNPNGRISIVIVKYKEVCPMMQTVYTVQVRSKPSLEVLITN 118

QY 121 LNPGETTYRITKVAENSAGIGVSDPFLFOTAESAPGVKVNLTVEAYNASAVKLIWYLPQ 180
Db 119 LNPGETTYRITKVAENSAGIGVSDPFLFOTAESAPGVKVDFTGEAVPFS--KLWYTSAT 177

QY 181 PNGKITSPKISVKHARSGIVVKDVSIRVEDILITGKL--ECNENSESFLWSTASPSPLGR 239
Db 178 KK-KITSPKISVKHNRSGIVVKEVSIRVECILSASLPLHCNENSESFLWSTASPSPLGR 236

QY 240 VTTPSRITTHSSSTLTQNEISSVVKKEPIISFVWTHLRPYTYLFEVSAATTEAGYIDSTIYR 299
Db 237 VTTPSRITTHSSSTLTQNEISSV--KEPIISFVWTHLRPYTYLFEVSAATTEAGYIDSTIYR 295

QY 300 TPESVPEGPPQNCVNTGNITGKSFSILWDPPTIVTGKFSYRVLYGPS--GRILDNSTKDLK 358
Db 296 TPESVPEGPPQNCVNTGNITGKSFSILWDPPTIVTGKFSYRVLYGPSAGRIILDNSTKDLK 355

QY 359 FAFNTLTPFTMYDVYIAAETSAGTGPKSGNISVFTPPDVPFGAVFDLQLAEVESTQVRIWK 418
Db 356 FAFNTLTPFTMYDVYIAAETSAGTGPKSGNISVFTPPDVPFGAVFDLQLAEVESTQVRIWK 415

QY 419 KPRQNGIINQYRVKVLVPETGIIILENTLLTGNEYINDPMAPEVINIVBPMVGLYEGSA 478
Db 416 KPRQNGIINQYRVKVLVPETGIIILENTLLTGNE--INDPMAPEVINIVQPMVGLYEGSA 474

QY 479 EMSSDLHSIATPIYNSHPDKPPARNRAEDQTSPPVVTTRNOYITDIAEOLSVYVRLVLP 538
Db 475 EMSSDLHSIATPIYNSHPDKPPARNRAEDQTSPPVVTTRNQYITDIAEOLSVYVRLRLR 534

QY 539 FTEHMSVSAFTIMGEGPPTVLSVTRTQVPSIKIINYK--NISSSILLWDPPEYN 596
Db 535 FWAETMGFSRYTIMSS-----ASRDLNLTSPGLSAQNFVTHVITTEVFLHWDPD--P 586

QY 597 GKITHYTIYAMELDT-NRAFOITITDN-SFLITGLKKTYYKQVRAASTHDESSISEEN 654
Db 587 VFFHYLITILDVENQSKSIILRLNLSLSLVILGLKKTYYKQVRAASTHVGESSISEEN 646

QY 655 DIFVTSDEPSSPDVEVIDVTADEIRLKWSPPEKNGIILAYEVLKNTDLYMKNT 714
Db 647 DIFVTSDEPSSPDVEVIDVTADEIRLKWSPPEKNGIILAYEVLKNTDLYMKNT 706

QY 715 STTDIILNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSPAENITYKNISG 774
Db 707 STTDIILNLRPHLTLYNISVRSYTRFSGHGNQVSSLLSVRTSETVPDSPAENITYKNISG 766

QY 775 EIELSFLPPSSPNGIILKTYIYKRSNGNEERTINTSLTQNIKVLKKTQYIIEVSAST 834
Db 767 EIELSFLPPSSPNGIILKTYIYKRSNGNEERTINTSLTQNIKVLKKTQYIIEVSAST 826

QY 835 LKGEGRVSAPISILTEEDAPSPDQFSVKQISGVTVKLSWOPPEKNGIILYYTVVYN 894
Db 827 LKGEGRVSAPISILTEEDAPSPDQFSVKQISGVTVKLSWOPPEKNGIILYYTVVYN 885

QY 895 RSSLKTINVTTLSLESLDLNVNVEYSAYVTASTRFGDGKGTGNSNIISFQTEGAPSDPPK 954
Db 886 RSSLKTINVTTLSLESLDLNVNVEYSAYVTASTRFGDGKGTGNSNIISFQTEG-PSDPPK 944

QY 955 VTYANLSSSSIIILFWTPPSKNGIILQYYSVYVRYNTSGTFMQNFTLHETLNDNDNVTSTI 1014
Db 945 VTYANLSSSSIIILFWTPPSKNGIILQYYSVYVRYNTSGTFMQNFTLHETVNDNDNVTSTI 1004


```
QY 1015 IDKLTIFSYVYFWLTASTSVGNKSSDIEVVYTDQDIPBGFVGNLTYSISSTAINVSW 1074
DB 1005 IDKLTIFSYVYFWLTASTSVGNKSSDIEVVYTDQDQVPEGFVGNLTYSISSTAINVSW 1064
QY 1075 VPPAQPNGLVFFVYVLSILQOTPRHVPPLVTVYERSIYFONLEKYTDYILKITPSTBKGFS 1134
DB 1065 VPPAQPNGLVFFVYVLSILQOTPRHVPPLVTVYERSIYFONLEKYTDYILKITPSTBKGFS 1124
QY 1135 DYTATQALYIKTEBEDVPETSPINTFKNLSSTSVLLSWDPVPKNGAILIYVDLTLOQPNEN 1194
DB 1125 DYTATQALYIKTEBEDVPETSPINTFKNLSSTSVLLSWDPVPKNGAILIYVDLTLOQPNEN 1184
QY 1195 YSFLTSDNVIILBELSPFTLYSFAAARTKGLGRSSILFFYTDQVPLAPQNLTLINC 1254
DB 1185 YSFLTSDNVIILBELSPFTLYSFAAARTKGLGRSSILFFYTDQVPLAPQNLTLINC 1244
QY 1255 TSDPFWLKWSPSPGCIKVKVYFKIHEHETDIYKNIISGFKTEAKLVGLEPVSYSIR 1314
DB 1245 TSDPFWLKWSPSPGCIKVKVYFKIHEHETDIYKNIISGFKTEAKLVGLEPVSYSIR 1304
QY 1315 VSAFTKVGNGQPSNVVVKFTTQESVDPVQNMOCMATSWQSVLVKWDPPKANGIITQM 1374
DB 1305 VSAFTKVGNGQPSNVVVKFTTQESVDPVQNMOCMATSWQSVLVKWDPPKANGIITQM 1364
QY 1375 VTVERNSTKVSPODHMYTPIKLLANTSYVFKVRASTAGGEDSTCHVSTLPETVPSVPT 1434
DB 1365 VTVERNSTKVSPODHMYTPIKLLANTSYVFKVRASTAGGEDSTCHVSTLPETVPSVPT 1424
QY 1435 NTAFSVDQSTATLTWIRPDTILGYFQNYKIITQLRAQCKWESECEVEYOKIQLYEA 1494
DB 1425 NTAFSVDQSTATLTWIRPDTILGYFQNYKIITQLRAQCKWESECEVEYOKIQLYEA 1484
QY 1495 HLTEETVYGLKFRWYRFQVAASTNAGYNASNIWISTKLPDGPDPGPPENHVHVAATSPS 1554
DB 1485 HLTEETVYGLKFRWYRFQVAASTNAGYNASNIWISTKLPDGPDPGPPENHVHVAATSPS 1544
QY 1555 ISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTEIKDLBIFRYSVVITA 1614
DB 1545 ISISWSEPAVITGPTCYLIDVKSVDNDEFNISPISKEENKTEIKDLBIFRYSVVITA 1604
QY 1615 FTGNTSAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPSPQNGNI 1674
DB 1605 FTGNTSAAVVEGKSSAEMIVTTLESAPKDPNNMTFQKIPDEVTKFQLTFLPSPQNGNI 1664
QY 1675 QYVQALVYREDDPTAQVHNLISIIOKTNTFVIAMLEGLKGHTYNI SVYAVNSAGAGPKV 1734
DB 1665 QYVQALVYREDDPTAQVHNLISIIOKTNTFVIAMLEGLKGHTYNI SVYAVNSAGAGPKV 1724
QY 1735 PMRITMDIKAPARPKTKPTPIYDATGKLVSTTTIIRMPICYSDDHGPIKNVQVLATE 1794
DB 1725 PMRITMDIKAPARPKTKPTPIYDATGKLVSTTTIIRMPICYSDDHGPIKNVQVLATE 1784
QY 1795 TGAQHDGNTWKYDAVFNKARPYFTNEGPPNPCTEGKTKPSGNEBIYIIGADNACMIBG 1854
DB 1785 TGAQHDGNTWKYDAVFNKARPYFTNEGPPNPCTEGKTKPSGNEBIYIIGADNACMIBG 1844
QY 1855 NEDKICNGLPKPKQVLFKFRATNIMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1914
DB 1845 NEDKICNGLPKPKQVLFKFRATNIMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCI 1904
QY 1915 LSIILLGTAIFAPARIQOKBGGTYSPODABIIDTKLKLQDITVADIELKDERLTRPI 1974
DB 1905 LSIILLGTAIFAPARIQOKBGGTYSPODABIIDTKLKLQDITVADIELKDERLTRPI 1964
QY 1975 SKKSPLOHVEELCTNNLKFQEFSEL PKFLODLSSTADLDPNRAKRPNNI KPNNNN 2034
DB 1965 SKKSPLOHVEELCTNNLKFQEFSEL PKFLODLSSTADLDPNRAKRPNNI KPN 2023
QY 2035 RVKLLADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRWVWETRAKTLVMLT 2094
DB 2024 RVKLLADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFWRWVWETRAKTLVMLT 2083
QY 2095 QCFEKGRIRKCHQWPEDNKPVTVFGDIVITIKLMEDVQIDWTIRDLKIERHGDCMTVRQCN 2154
```

```
DB 2084 QCFEKGRIRKCHQWPEDNKPVTVFGDIVITIKLMEDVQIDWTIRDLKIERHGDCMTVRQCN 2143
QY 2155 FTAWPEHGVPENSAPLIHFVKLVRASTRAHDTTTPMIVHCSAGVGRGTGVFIALDHLTQHIND 2214
DB 2144 FTAWPEHGVPENSAPLIHFVKLVRASTRAHDTTTPMIVHCSAGVGRGTGVFIALDHLTQHIND 2203
QY 2215 HDPVDIYGLVABLRSEKMCWQNLQAYIFLHQICILDLNKGNSQBPICFVNTYSALQKQMS 2274
DB 2204 HDPVDIYGLVABLRSEKMCWQNLQAYIFLHQICILDLNKGNSQBPICFVNTYSALQKQMS 2263
QY 2275 LDAME-GDVELEWEETTM 2291
DB 2264 LDAMEGDVELEWEETTM 2281
```

RESULT 8

```
US-09-822-871-4
; Sequence 4, Application US/09822871
; Publication No. US20030099942A1
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS.
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001219
; CURRENT APPLICATION NUMBER: US/09/822,871
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2301
; TYPE: .PRT
; ORGANISM: Rat
US-09-822-871-4
```

Query Match 87.9%; Score 10536; DB 10; Length 2301;

Best Local Similarity 85.9%; Pred. No. 0;

Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

```
QY 1 MDPLIIFLLFLLTGTSQVDSNVVPGTRYDITISSIS-TTYSPTVTRIVTNVTKPGPP 59
DB 1 MDHFHSFLLIIGTSQVDSNVVPGTRYDITISSIS-TTYSPTVTRIVTNVTKPGPP 60
QY 60 VFLAGERSGAGILLISWNTPPNPNNGRIISYIVKYEVCPPMOTVYTVQVSKPDSLEVLIT 119
DB 61 VFLAGERSGAGILLISWNTPPNPNNGRIISYIVKYEVCPPMOTVYTVQVSKPDSLEVLIT 120
QY 120 NLNPGTTTTEIKVAENNSAGIGVFSDDPFLPQTAEAPGKVVNLTVREYNASAVKLIWYLR 179
DB 121 NLNPGTTTTEIKVAENNSAGIGVFSDDPFLPQTAEAPGKVVNLTVREYNASAVKLIWYLR 180
QY 180 QPNKITSFKISVKHARSIGIVVKQVSRVEDILTGLPECNSENSEFLWSTASPSPLGR 239
DB 181 QPNKITSFKISVKHARSIGIVVKQVSRVEDILTGLPECNSENSEFLWSTASPSPLGR 240
QY 240 VTPPSRTTSSSTLQNEISSVWKEPISFVVTTHLPYTYLFEVSAATTEAGYIDSTIVR 299
DB 241 VTPVTRTQSSSTAARSKISSVWKEPISFVVTTHLPYTYLFEVSAATTEAGYIDSTIVR 300
QY 300 TPESVPEGPPQNCVCTGNITGKSFSLMDPPTTIVTKFSYRVELYGPSGRILONSTKOLKP 359
DB 301 TPESVPEGPPQNCVCTGNITGKSFSLMDPPTTIVTKFSYRVELYGPSGRILONSTKOLKP 360
QY 360 AFTNLTPMTVDVYTAETSAGTSPKSNISVTPPDVPGVADQLQAEVSTQVRLTWK 419
DB 361 AFTNLTPMTVDVYTAETSAGTSPKSNISVTPPDVPGVADQLQAEVSTQVRLTWK 420
QY 420 PRQNGIINQVRKVLVPEPTEGIIENLTLLTGNBNYINDPMAPEIVNIPEPVMVGLYEGSAB 479
DB 421 PRQNGIINQVRKVLVPEPTEGIIENLTLLTGNBNYINDPMAPEIVNIPEPVMVGLYEGSAB 480
QY 480 MSSDLHSLATFTYNSHPDKNFPARNABEDQTSFVVVTRNQYITDIAAEQLSVIRRLVFP 539
```

```
Db      481  MSDDLSPASFIYNSHPNDPFASTRABEQSPVVTTRNQYMTDITABQLSVVRRLVFP 540
Qy      540  TTHMTLSVSAFTMGEPPTVLSVRTRQVPSISKIINYKNISSSSILLWDPPEYPNGKI 599
Db      541  THTHTSVSAFTMGEPPTVLSVRTRQVPSIQIINYKNISSSSILLWDPPEYPNGKI 600
Qy      600  THYTIYAMELDTNRAFIQTIDNSFLITGLKYYTKYKXKVAASATHDGSSLSSEENDIFVR 659
Db      601  THYTIYATELDTNRAFOFTVDNSFLITGLKYYTKYKXKVAASATHVGSSLSSEENDIFVR 660
Qy      660  TSEDEPSSQDVEVDVTADEIRLKWSPPEKNGIIIAYEVLVYKNIIDTLVMKNTSTTDI 719
Db      661  TPEDEPSSQDQVQVGVSPSELRLKWSPEKPNGIIIIAYEVLVYQWADTLFVKNTSTTDI 720
Qy      720  ILLENIRPHLTNYISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITVKNLSSGEIELS 779
Db      721  IISDLKPYTLNINSIRSLRGLGHGNQSSLLSVRTSETVPDSAPENITVKNLSSGEIELS 780
Qy      780  FLPPSPNGIIRKYYTILKRSNGNBERTINTTSLQNIKVLKXYQYIIIEVSASTLKGBG 839
Db      781  FLPPSPNGIIRKYYTILKRSNSHEARTINTTSLQTIIGLKKYTHYVIEVSASTLKGBG 840
Qy      840  VRSAPISILTEBAPDPSPPQDSVQLSGVTVKLSWQPPLEPNGLIYYTVVWNRSLK 899
Db      841  IRSRPSILTEBAPDPSPPQDSVQLSGVTVKLSWQPPLEPNGLIYYTVVWMDKSLR 900
Qy      900  TINVETSELSGLDYNVEYSAYVTASTRFQGGKTSNLIISQTPREGAPSDPKOVYYAN 959
Db      901  AINATEASLVLDLDYNDYGACVTASTRFGDGNARSSIIINFRTEGEPSPDNDVHYVN 960
Qy      960  LSSSSILFWTPPPKNGIIQYYSVYRNTSGTFMQNFHETLHETDNDMTVTIIDKLT 1019
Db      961  LSSSSILFWTPPVKNGIIQYYSVYQNTSGTFVQNFLLQVTKESDNVTVSARIYLA 1020
Qy      1020  IFSYTFWLTASTSVGNKSSDIIIEVYTDQDPIEGFVGNLTYESISSSTAINVSWPPAQ 1079
Db      1021  IFSYTFWLTASTSVGNKSSDIIIEVYTDQDPIEGFVGNLTYESISSSTAIHVSWEPPSQ 1080
Qy      1080  PNGLVFYVYLLILOQT-PRHVRPLVYERSIYFDNLEKYYTIIKITSTEGSDTYT 1138
Db      1081  PNGLVFYVYLLSQSPPHMTPLVLYENSIDFDLEKYTDYIKFITSTEGKSETY 1140
Qy      1139  AOLYKITEEDVPETSPINFTKNLSTSVLLSWDPPVKNGAISYDLTLQCPNENYSFI 1198
Db      1141  TQHLKITEEDVPETSPINFTKNLSTSVLLSWDPPVKNGAILGVHLTLQCPHANHTFV 1200
Qy      1199  TSDNYIILELSPFTLYSFAAARTKGLSPSIIFFYTDVSVPLAPPQNLTILNCTSD 1258
Db      1201  TSGNHI VLEELSPFTLYSFAAARTMKGILGSSIIFFYTDVSVPLAPPQNLTILNCTSD 1260
Qy      1259  VMLKWSPPSLPGGIVKVKYSEKHEHETDIYKNIISGFKTEAKLVGLEPVSVYSIRVSF 1318
Db      1261  VMLTWSPPSLPGGIVKVKYSEKHEHETDIYKNIISGLQTDKLEGLFVSVYSVSFA 1320
Qy      1319  TKVGNQNSNVKFTTQSVDPVQVNMOCMATSQSVLKVWDPPKKANGIITQYVMTVE 1378
Db      1321  TKVGNQNSNVVEFTTQSVPEAVRNICVARDMQSVSRWDPDPKKTNGIILHMTVIG 1380
Qy      1379  RNSTKVSPODHMYTFIKLLANTSYVPKVRASTSAGEDESTCHVSTLPETVPSVTNIAF 1438
Db      1381  GNSTKVSPPDPTVFTFKLLPNTSYVEVRASTSGESNRCDSITLPETVPSAPTNVAF 1440
Qy      1439  SDVQSTSATLTWRPDTILGYFQNYKITTLQRAQKKEWSEBCEVYKIQIYEAHLTE 1498
Db      1441  SNVQSTSATLTWKPTDITFYQNYKITTLQRAQKKEWSEBCEVYKIQIYEAHQVTE 1500
Qy      1499  ETVYGLKKRWRFQVAASNAGYGNASNIWSTKLPGPPDPPENHVAVTSPPSIS 1558
Db      1501  ETVHGLKKRWRFQVAASNAGYGNASNIWSTKLPGPPDPPENHVAVTSPPGNIS 1560
Qy      1559  WSEPAVITGPTCYLIDVKSVDNDEFNISFKSENEENKTIIEIKDLEIFTRYSVVITAFGN 1618
```

```
Db      1561  WSEPAVITGPTCYLIDVKSVDNDEFNISFKSENEENKTIIEINLEVFTRYSVVITAFVGN 1620
Qy      1619  ISAAAYVEGKSSAEMIVTTLESAPKPPNNMTFQKIPDEVTQFQLTFLPSPQNGNIQVYQ 1678
Db      1621  VSRAITDCKSSAEVITTTLESVPKPPNNMTFQKIPDEVTQFQLTFLPSPQNGNIRVYQ 1680
Qy      1679  ALVYREDDPTAVQHNLSIIOKTNTFVIAMLEGLKGGHTYNIISVAVNSAGAGPKVPMRI 1738
Db      1681  ALVYREDDPTAVQHNFSIIOKTNTSIIAMLEGLKGGHTYNIISVAVNSAGAGPKVQMRI 1740
Qy      1739  TMDIKAPARPKTPIYDATCKLLVTTTIRMPICVYSDHGHPIKNVQVLAETGAQ 1798
Db      1741  TMDIKAPARPKSPIPIRDATCKLLVTTTIRMPICVYNDHGHPIRNVQVLAETGAQ 1800
Qy      1799  HDGNTYKWDAYFNKARPVFTNEGPNPCFEGKTKFSGNEEIIYIIGADNACMIPGNEBK 1858
Db      1801  QDGNVTKWDAYFNKARPVFTNEGPNPCFEGKTKFSGNEEIIYIIGADNACMIPGNEBK 1860
Qy      1859  ICNGPLKPKQYLFKFRATNIMGQFSDSDYSPVTKLGEGLSERTVEILSVTLCILSII 1918
Db      1861  ICNGPLKPKQYLFKFRATNVMGQFTDSEYSDPIKTLGEGLSERTVEILSVTLCILSII 1920
Qy      1919  LLGTAIFAFARIRQKQEGGYSPODASIIDTKLKDQILITVADLELKDRLTR----- 1972
Db      1921  LLGTAIFAFARIRQKQEGGYSPRDASIIDTKLKDQILITVADLELKDRLTRLLSVRK 1980
Qy      1973  ---PIKSKSFLOHVEELCTNNLKFQEFSELKFLQDLSSTDADLPNNRAKNRFPNIKP 2029
Db      1981  STKPIKSKSFLOHVEELCTNNLKFQEFSELKFLQDLSSTDADLPNNRAKNRFPNIKP 2040
Qy      2030  YNNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTGVDFWVWVETRAKT 2089
Db      2041  Y-NNNRVKLIADVSLPGSDYINASYISGYLCPNEFIATQGPLGTGVDFWVWVETRAKT 2099
Qy      2090  LVMLTQCEKGRIRCHOVWPDENKVPYVGDIVITKLMEDVQIDMTIRDLKIERHGDGMT 2149
Db      2100  LVMLTQCEKGRIRCHOVWPDENKVPYVGDIVITKLMEDVQIDMTIRDLKIERHGDGMT 2159
Qy      2150  VQRCNFTWAPHEGVSPENSAPLHFKVLRASRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2209
Db      2160  VQRCNFTWAPHEGVSPENTPLHFKVLRASRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2219
Qy      2210  QHINDHDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLNKGNSQPICFVNYSA 2269
Db      2220  QHINNHDVFDIYGLVAELSERMCMVQNLAQYIFLHQICILDLNKGNSQPICFVNYSA 2279
Qy      2270  QKMDSLDAMEGDVELEWSETTM 2291
Db      2280  QKMDSLDAMEGDVELEWSETTM 2301
```

RESULT 9

```
US-10-673-885-4
; Sequence 4, Application US/10673885
; Publication No. US20040081644A1
; GENERAL INFORMATION:
; APPLICANT: BRASLEY, Ellen et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: AND USES THEREOF
; FILE REFERENCE: CL001219DIV
; CURRENT APPLICATION NUMBER: US/10/673,885
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 09/822,871
; PRIOR FILING DATE: 04-02-2001
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2301
; TYPE: PRT
; ORGANISM: Rat
US-10-673-885-4
```

Query Match 87.9%; Score 10536; DB 15; Length 2301;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

QY 1 MDPLIIFLLFICTSETQVDSNVVPGTRDYDITISS-ITTSPTVRIWNTVTKGPP 59
DB 1 MDHPSFLFLICTSESQVDSVSSFDGTGYDITLSSVSATTSYSPVSRILTATVTKGPP 60

QY 60 VFLAGRVGSAGILLSNWTPNPNNGRIISYVYKKEVCWPMQVYTVQVRSKPSLEVLIT 119
DB 61 VFLAGRVGSAGILLSNWTPNPNNGRIISYVYKKEVCWPMQVYTVQVRSKPSLEVLIT 120

QY 120 NLNPGTYEILKVAENSAGIGVSDPFLPOTASAPGKVVNLTVAYNASAVKLWYLPR 179
DB 121 NLNPGTYEILKVAENNAGIGVSDPFLPOTASAPGKVVNLTVAYNASAVKLWYLPR 180

QY 180 QPNKITSFKISVKHARSIGVVDVSRVEDIITLGLPECNENSESFLMSTASPSPTLGR 239
DB 181 QPNKITSFKISVKHARSIGVVDVSRVEDIITLGLPECNENSESFLMSTASPSPTLGR 240

QY 240 VTPSRTHSSSTLTQNEISSVWKEPISFVVTHLRPYTYLFEVSAATTEAGYIDSTIVR 299
DB 241 VTPSRTHSSSTLTQNEISSVWKEPISFVVTHLRPYTYLFEVSAATTEAGYIDSTIVR 300

QY 300 TPESVPEGPPQNCVNTGKSFESILWDPPTIIVTGFSYRVELYGSGRILDNSTKDLKF 359
DB 301 TPESVPEGPPQNCVNTGKSFESILWDPPTIIVTGFSYRVELYGSGRILDNSTKDLKF 360

QY 360 APTNLPFTWYDVYIAAETSAGTSPKSNISVFTPPDVGAFDLOLAEVESQVRIWTKK 419
DB 361 APTNLPFTWYDVYIAAETSAGTSPKSNISVFTPPDVGAFDLOLAEVESQVRIWTKK 420

QY 420 PRQNGIINQYRVKVLVPEGTGIIILENTLLTGNNEYINDPMAPEIVNIBPVMGLYEGSAB 479
DB 421 PRQNGIINQYRVKVLVPEGTGIIILENTLLTGNNEYINDPMAPEIVNIBPVMGLYEGSAB 480

QY 480 MSSDLHSLATFYNHSHDKNPPARNRAEDOTSQVTRNQYITDIAEOLSVIRRLVFP 539
DB 481 MSSDLHSLATFYNHSHDKNPPARNRAEDOTSQVTRNQYITDIAEOLSVIRRLVFP 540

QY 540 TEHMIVSAPFTINGEGPPVTLVSRTRQVPSSTIKIINYKNISSSSILLYWDPEYNGKI 599
DB 541 TEHMIVSAPFTINGEGPPVTLVSRTRQVPSSTIKIINYKNISSSSILLYWDPEYNGKI 600

QY 600 THYTIYAMELDTNRAFQITTDINSFLITGLKTKYTKMRVAASHTDGSLSSENDIFVR 659
DB 601 THYTIYAMELDTNRAFQITTDINSFLITGLKTKYTKMRVAASHTDGSLSSENDIFVR 660

QY 660 TSEDEPESQDVVEIDVTADIEIRLKWSPPEKNGIIIAEVLVYKNIIDLYMKNTSTTDI 719
DB 661 TSEDEPESQDVVEIDVTADIEIRLKWSPPEKNGIIIAEVLVYKNIIDLYMKNTSTTDI 720

QY 720 ILRLNLPHTLYNISVRSYTRPGHGNQVSSLSVRTSETVPDSAPENITTKNISSGIEIS 779
DB 721 ILRLNLPHTLYNISVRSYTRPGHGNQVSSLSVRTSETVPDSAPENITTKNISSGIEIS 780

QY 780 FLPPSPNGIISKYTYILKRSNGNEBRTINTSLTONIKVLKYTOYIIIEVSASTLKGRG 839
DB 781 FLPPSPNGIISKYTYILKRSNGNEBRTINTSLTONIKVLKYTOYIIIEVSASTLKGRG 840

QY 840 VRSAPISILTEEDAPSPQDPSVKOLSGVTVKLSWQPLEPENGIIILYTVYVWNRSLK 899
DB 841 VRSAPISILTEEDAPSPQDPSVKOLSGVTVKLSWQPLEPENGIIILYTVYVWNRSLK 900

QY 900 TINVTETSELSLDNVEYSAVTASTRFGDKTGSNIIISQTPGAPSDPPKDVYAN 959
DB 901 TINVTETSELSLDNVEYSAVTASTRFGDKTGSNIIISQTPGAPSDPPKDVYAN 960

QY 960 LSSSSIIILFWTPPSKENGIIQYVSVYRYNTSGTFMONTLHETLNDPDMVTSTIIDKLT 1019
DB 961 LSSSSIIILFWTPPSKENGIIQYVSVYRYNTSGTFMONTLHETLNDPDMVTSTIIDKLT 1020

QY 1020 IFSYTFWLTASTSVGNKSSDIEVYTDQDIPEGVGNLTVESISSTAINVSWVPPAQ 1079

DB 1021 IFSYTFWLTASTSVGNKSSDIEVYTDQDIPEGVGNLTVESISSTAINVSWVPPAQ 1080

QY 1080 PNLGVFYVYSLILQOT-PRHVRPPLVYTERSYFONLEKYTYILKITESTKGFSDTYT 1138
DB 1081 PNLGVFYVYSLILQOT-PRHVRPPLVYTERSYFONLEKYTYILKITESTKGFSDTYT 1140

QY 1139 AQIYKTEEDVDPETSPIINTFKNLSTSVLLSWDPKPKNGALISDYLITQGNENYSP 1198
DB 1141 AQIYKTEEDVDPETSPIINTFKNLSTSVLLSWDPKPKNGALISDYLITQGNENYSP 1200

QY 1199 TSDNYIILEELSPFTLYSPFAAARTKGLGSPSSILFFYTDSEVPLAPPQNLTLINCTSD 1258
DB 1201 TSDNYIILEELSPFTLYSPFAAARTKGLGSPSSILFFYTDSEVPLAPPQNLTLINCTSD 1260

QY 1259 VMLKSPSPPLGGIVKVSFKIHEHETDIYIKNISGFKTEAKLVGLBVSIVSYRVSF 1318
DB 1261 VMLKSPSPPLGGIVKVSFKIHEHETDIYIKNISGFKTEAKLVGLBVSIVSYRVSF 1320

QY 1319 TKVGNQNSNVVKTFTQESVDDVQNMCMATSWQSVLVKWDPPKANGIITQYVMTVE 1378
DB 1321 TKVGNQNSNVVKTFTQESVDDVQNMCMATSWQSVLVKWDPPKANGIITQYVMTVE 1380

QY 1379 RNSTKVSPODHMYTFIKLLANTSYVFKVRASTAGDESDTCHVSTLPTVSPVPTNIAF 1438
DB 1381 RNSTKVSPODHMYTFIKLLANTSYVFKVRASTAGDESDTCHVSTLPTVSPVPTNIAF 1440

QY 1439 SDVQSTSATLWIRDDTLIGYFQNYKITTQLRAQKKEWSEBECVEYQIKVLYEAHLTE 1498
DB 1441 SDVQSTSATLWIRDDTLIGYFQNYKITTQLRAQKKEWSEBECVEYQIKVLYEAHLTE 1500

QY 1499 ETYVGLKPRWTRFQVAASTNAGYGNASNNISTKTLPGPDGPPENHVHVSATSPFISIS 1558
DB 1501 ETYVGLKPRWTRFQVAASTNAGYGNASNNISTKTLPGPDGPPENHVHVSATSPFISIS 1560

QY 1559 WSEPAVITGPTCYLIDVKSVDNDENISFKIKNENKTIIEIKDLEIFTSYVITAFVGN 1618
DB 1561 WSEPAVITGPTCYLIDVKSVDNDENISFKIKNENKTIIEIKDLEIFTSYVITAFVGN 1620

QY 1619 ISAAVVEGKSSAEMITVTTLESAPKDPNNMTFKIPDEVTQFQTLPLPSQPNNGINQVQ 1678
DB 1621 ISAAVVEGKSSAEMITVTTLESAPKDPNNMTFKIPDEVTQFQTLPLPSQPNNGINQVQ 1680

QY 1679 ALVYREDPPTAVQIHNLSIIQKNTFVIAMLEGLKGHTYNTSVYAVNSAGAPKVPMRI 1738
DB 1681 ALVYREDPPTAVQIHNLSIIQKNTFVIAMLEGLKGHTYNTSVYAVNSAGAPKVPMRI 1740

QY 1739 TWDIIPAPKPKTPPIVDATGKLLVSTTIIRMPICYSDHGPICKVQVLAETGAQ 1798
DB 1741 TWDIIPAPKPKTPPIVDATGKLLVSTTIIRMPICYSDHGPICKVQVLAETGAQ 1800

QY 1799 HDGNTWKYDAVFNKARPVFTNEGPNPCTSGKTKFSGNBEIYIIGADNACMIPGNEBK 1858
DB 1801 HDGNTWKYDAVFNKARPVFTNEGPNPCTSGKTKFSGNBEIYIIGADNACMIPGNEBK 1860

QY 1859 ICNGLPKPKQVLFKFRATNIMGQFSDSYSPVTKLGEGLSERTVEIILSVTLCLISII 1918
DB 1861 ICNGLPKPKQVLFKFRATNIMGQFSDSYSPVTKLGEGLSERTVEIILSVTLCLISII 1920

QY 1919 LIGTAIPAPARIQKQKGGTYSPODAEIIDTKLQDLITVADILEKDERLTR----- 1972
DB 1921 LIGTAIPAPARIQKQKGGTYSPODAEIIDTKLQDLITVADILEKDERLTR----- 1980

QY 1973 ---PISSKSFLOHVELCTNNLKFQEFSELPKFLQDLSSTADLPNNRKNRPNNIKP 2029
DB 1981 ---PISSKSFLOHVELCTNNLKFQEFSELPKFLQDLSSTADLPNNRKNRPNNIKP 2040

QY 2030 YNNNRVXKLADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFRWMWETRAKT 2089
DB 2041 YNNNRVXKLADASVPGSDYINASYISGYLCPNEFIATQGPLGTGDFRWMWETRAKT 2099

QY 2090 LVMLTQCCEKGRIRCHQWPEDNKPVTVFGDITVITKLMEDVQIDMTIRDLKIERHGDCMT 2149

Db 2100 LVLMTQCFKGRIRCHQYWPEDNKPVTVFGDIVITKLMEDIOIDWTIRDLKIERHGDCMT 2159
QY 2150 VRQCNFTAMPHEGVPENSAPLIHFVKLVRAASRAHDTTPMI VHCASAGVGRGTGVFIADHLT 2209
Db 2160 VRQCNFTGHPHGVPEHTTFLHFVKLVTSRAHDTTPMVHCASAGVGRGTGVFIADHLT 2219
QY 2210 QHNDHDFVDIYGLVAELSERCMQVONLAQYIFLHQICILDLLSNKGSNQPCFVNYNSAL 2269
Db 2220 QHNDHDFVDIYGLVAELSERCMQVONLAQYIFLHQICILDLLSNKGSNQPCFVNYNSAL 2279
QY 2270 QKMSLDAMEGDVELEWEETM 2291
Db 2280 QKMSLDAMEGDVELEWEETM 2301

RESULT 10

US-10-087-684-37
; Sequence 37, Application US/10087684
; Publication No. US20040029116A1

GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R.
; APPLICANT: MacDougall, John R.
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David J.
; APPLICANT: Grose, William M.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie, J.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Basha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 37
; LENGTH: 2302
; TYPE: PRT
; ORGANISM: Rattus norvegicus

US-10-087-684-37

Query Match 87.9%; Score 10536; DB 15; Length 2302;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;
QY 1 MDPLIFLLFIGTSETQDVSNVPGTRYDITISSIS-TTYSPTVRIIVTNVTKPGPP 59

Db 2 MDLHPSFLFLIGTSESQDVSSSPDGTGYDITLSSVSATTSYSSPVSTLATNVTKPGPP 61
QY 60 VFLAGERVGSAGILLSWTNPNGRIISYVKYKEVCPWMQTVTVTVQVRSKPDSDLEVLIT 119
Db 62 VFLAGERVGSAGILLSWTNPNGRIISYVKYKEVCPWMQTVTVTVQVRSKPDSDLEVLIT 121
QY 120 NLNPGTYYEIKVAAENSAGIVFSDPFLFQTAESAAGKVNLTVEAYNASAVKLLIWLPR 179
Db 122 NLNPGTYYEIKVAAENAGIVFSDPFLFQTAESAAGKVNLTVEALNYSVNLIIWLPR 181
QY 180 QNGKITSFKISVKHARSIGIVKOVSIKVEDILITKGLPECNENSSFTLWSTASPSPTLGR 239
Db 182 QNGKITSFKISVKHARSIGIVKOVSLVEDILITKGLPECNENSSFTLWSTASPSPTLGR 241
QY 240 VTPPSRTTHSSSTLTQNEISSVWKEPISFVVTHTLHPYTYTLFVESAATTEAGYIDSTIVR 299
Db 242 VTPVTRTTQSSSTAARSKISSVWKEPISFVVTHTLHPYTYTLFVESAATTEAGYIDSTIVR 301
QY 300 TPESVPEGPQNCVGTGKSFILMDPPITVTGKFSYRVELYGPGRIIDNSTKDLKF 359
Db 302 TPESVPEGPQNCVGTGKSFILMDPPITVTGKFSYRVELYGPGRIIDNSTKDLKF 361
QY 360 APTNLTPFTMDVYIAAETSAGTGPKNSIVFTPPDPVGAFLDLQAEVSTQVTRITWK 419
Db 362 APTNLTPFTMDVYIAAETSAGTGPKNSIVFTPPDPVGAFLDLQAEVSTQVTRITWK 421
QY 420 PROPNGIINOYRKVLVPETGILLENLTLLTGNNVINDMPAIEVINIYVPMVGLVGSAB 479
Db 422 PROPNGIISOYRKVSVLETVGVLENTLLTGQDESINPMSPFEMNLVDPMLGTFEGSGE 481
QY 480 MSSDLHSLATFIYNHSHDPKFNPARNAEDQTSPPVTTNRQYITDIAAQLSIVIRRLVPF 539
Db 482 MSSDLHSPASFIYNHSHDPNDPASTRAEQSSPVTTENQYMTDITAEQLSVVVRRLVPF 541
QY 540 TEHMISVSAPFTIMGEGPPTVLSVRTRQVPSIKIINYNKNISSSILYWDPPPEPNKGI 599
Db 542 TEHTISVSAPFTIMGEGPPTVLTREQVPSIIQIINYNKNISSSILYWDPPPEPNKGI 601
QY 600 THYTIYAMELDTNRAFOITTDNSFLITGLKKYTKYKRVAASTHGDGSSISEENDIFVR 659
Db 602 THYTIYATELDTNRAFOITTDNSFLITGLKKYTKYKRVAASTHGDGSSISEENDIFVR 661
QY 660 TSEDEPSSPODEVIDVTADIEIRLKWPPKPNKGIITAYEVLKNIIDTLYMKNSTTDDI 719
Db 662 TPEDEPSSPODEVIDVTADIEIRLKWPPKPNKGIITAYEVLKNIIDTLYMKNSTTDDI 721
QY 720 ILRLNRLPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSEIELS 779
Db 722 IISDLKPVTLYNISRSYTRFGHGNQVSSLLSVRTSETVPSAPENITYKNISSEIELS 781
QY 780 FLPPSSPNGLIKKTYIYLRKSNNGEERTINTSLTONIKVLKKYTOYIIEVSASTLKGG 839
Db 782 FLPPSSPNGLIKKTYIYLRKSNNGEERTINTSLTONIKVLKKYTOYIIEVSASTLKGG 841
QY 840 VRSAPISILTEADAPDPSPQVSVKQLSGVTVKLSWQPPLEPNGLIYTYVVMNRSLK 899
Db 842 IIRSFISILTEADAPDPSPQVSVKQLSGVTVKLSWQPPLEPNGLIYTYVVMNRSLK 901
QY 900 TINVTETSLSDLYNVVEYSAVYTAFTRGDKTGSNIISPTQTEGAPSDPPKDVYAN 959
Db 902 AINATEASLSDLYNVVDYGCACVTAFTRGDKGNARSIINPRTEGEPSPDPNDVHYN 961
QY 960 LSSSSIIILFWTPPSKPNGLIQYISYVYRNTSGTFMONTLHETLNDFNMTVTIIDLKT 1019
Db 962 LSSSSIIILFWTPPSKPNGLIQYISYVYQNTSGTFMONTLHETLNDFNMTVTIIDLKT 1021
QY 1020 IFSYTYFWLTASTSVGNKSSDIEVYTDODIPEGFVGNLTYESISTAINVSVWPPAQ 1079
Db 1022 IFSYTYFWLTASTSVGNKSSDIEVYTDODIPEGFVGNLTYESISTAINVSVWPPAQ 1081
QY 1080 ENGLVYFVYSILQOT-PRHVRPPIVTVYERSIYFDNLEKTYDILKITPSTBKGSQDYTT 1138

Db 1082 PNGLVFYLLSLNQQSPRHMIPLVYIENSIDFDDLEKDYDIFKLTSTKGFSETVT 1141
Qy 1139 AQYIKTEEDVPETSPINTPKNLSTSVLLSWDPVPKPNCAIISYDLTLQGNENYSFI 1198
Db 1142 TQLHIKTEEDVPETSPINTPKNLSTSVLLSWDPVPKPNCAIISYDLTLQGNENYSFI 1201
Qy 1199 TSDNYILLESPTLYSFAAARTKGLGPSSILFYTDDESAPLAPPQNLTLINTYDSF 1258
Db 1202 TSGNHIVLESPTLYSFAAARTKGLGPSSILFYTDDESAPLAPPQNLTLINTYDSF 1261
Qy 1259 VMLKWSPLPGGIVKVSFKIHEHETDTIYKNSIGFKTEAKLVLEPVSTYSIRVSAP 1318
Db 1262 VMLTWSPLPGGIVKVSFKIHEHETDTIYKNSIGFKTEAKLVLEPVSTYSIRVSAP 1321
Qy 1319 TKVNGNQSNVVKFTQESVPDVQVQMCWATSWSVLVKWDPKPKANGIITQYMYTVE 1378
Db 1322 TKVNGNQSNVVKFTQESVPDVQVQMCWATSWSVLVKWDPKPKANGIITQYMYTVE 1381
Qy 1379 RSTKVSQDHYMTFFIKLLANTSYVFKVRASTAGDESDTCHVSTLPETVSPVNIAP 1438
Db 1382 GNSTKVSQDHYMTFFIKLLANTSYVFKVRASTAGDESDTCHVSTLPETVSPVNIAP 1441
Qy 1439 SDVQSTSATUTWRPDTILGYFQNYKITTLQRAQCKEWESECEVQKIYLAHLTE 1498
Db 1442 SNVQSTSATUTWRPDTILGYFQNYKITTLQRAQCKEWESECEVQKIYLAHLTE 1501
Qy 1499 ETYVGLKFRWYFQVAASTNAGYGNASWISFTLPGPPDGPENVHVVAATSPFISIS 1558
Db 1502 ETYVGLKFRWYFQVAASTNAGYGNASWISFTLPGPPDGPENVHVVAATSPFISIS 1561
Qy 1559 WSEPAVITGPTLYLIDVKSVDNDENFISFKSNEENKTIIEKDLFTFYSVITAFGN 1618
Db 1562 WSEPAVITGPTLYLIDVKSVDNDENFISFKSNEENKTIIEKDLFTFYSVITAFGN 1621
Qy 1619 ISAYVEGKSSAEVITLESAPKDPNNMTFOKIDEVTKQLTFLPPSQPNQIVQ 1678
Db 1622 VSRAYTDGSSAEVITLESAPKDPNNMTFOKIDEVTKQLTFLPPSQPNQIVQ 1681
Qy 1679 ALVYREDPTAVQIHNLSIIQKNTFVIAMLEGKGGHYNTSVYAVNSAGAGPKVPMRI 1738
Db 1682 ALVYREDPTAVQIHNLSIIQKNTFVIAMLEGKGGHYNTSVYAVNSAGAGPKVPMRI 1741
Qy 1739 TMDIKAPARKPTPIYDATGKLLVTSTITIRMPICYSDHGHGPKVQVLAETGAQ 1798
Db 1742 TMDIKAPARKPTPIYDATGKLLVTSTITIRMPICYSDHGHGPKVQVLAETGAQ 1801
Qy 1799 HDGNTKWDAYFNKARPFTNEGFNPPCTEGTKFSGNEHYIIGADNACMIPGNEBK 1858
Db 1802 QDGNVTKWDAYFNKARPFTNEGFNPPCTEGTKFSGNEHYIIGADNACMIPGNEBK 1861
Qy 1859 ICGNPLKPKQYLFKFRATNIMQFTDSYSDPVKTLGGLSERTVEIILSVTLCLISII 1918
Db 1862 ICGNPLKPKQYLFKFRATNIMQFTDSYSDPVKTLGGLSERTVEIILSVTLCLISII 1921
Qy 1919 LIGTAFAPARIQKQEGTYSPODAEIIDTKLKLQDITVADLEKDLERLTR----- 1972
Db 1922 LIGTAFAPARIQKQEGTYSPODAEIIDTKLKLQDITVADLEKDLERLTR----- 1981
Qy 1973 ----PISKKSFLQVHELCNNLKFQEPSELPKFLQDLSSTADLPNNRKNRPNIKP 2029
Db 1982 SIKPISKKSFLQVHELCNNLKFQEPSELPKFLQDLSSTADLPNNRKNRPNIKP 2041
Qy 2030 YNNNRVKLIADASVPGSDYINASYISGVLCPNEFIATQGLPGTGVDFRWWMETRKT 2089
Db 2042 Y-NNNRVKLIADASVPGSDYINASYISGVLCPNEFIATQGLPGTGVDFRWWMETRKT 2100
Qy 2090 LVMLTQCFEKGIRCHQYWPENKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCT 2149
Db 2101 LVMLTQCFEKGIRCHQYWPENKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCT 2160
Qy 2150 VRQCNFTAWPERGVNPENAPLIFHVKLVRASRAHDTTPMIVHCSAGVGTGVFIADHIT 2209
Db 2161 VRQCNFTAWPERGVNPENAPLIFHVKLVRASRAHDTTPMIVHCSAGVGTGVFIADHIT 2220

Qy 2210 QHINHDFVDIYGLVAELRSEKCMQVNLQAQIFLHQICILDLNKGSKNQPICFVNSAL 2269
Db 2221 QHINHDFVDIYGLVAELRSEKCMQVNLQAQIFLHQICILDLNKGSKNQPICFVNSAL 2280
Qy 2270 QKMSDLDAMEGDEVELEWETTM 2291
Db 2281 QKMSDLDAMEGDEVELEWETTM 2302

RESULT 11

US-10-218-779-37
; Sequence 37, Application US/10218779
; Publication No. US2004002922A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William
; APPLICANT: Alsobrook II, John
; APPLICANT: Lepley, Denise
; APPLICANT: Rieger, Daniel
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Spylek, Kimberly
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy Suresh
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar
; APPLICANT: Vernet, Corine
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Malyankar, Uriel
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles
; APPLICANT: Gangolli, Esha
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-214
; CURRENT APPLICATION NUMBER: US/10/218,779
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,-926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 37
; LENGTH: 2302
; TYPE: PRT
; ORGANISM: Rattus rattus
US-10-218-779-37

Query Match 87.9%; Score 10536; DB 15; Length 2302;
Best Local Similarity 85.9%; Pred. NO. 0;
Matches 1977; Conservative 162; Mismatches 151; Indels 12; Gaps 4;

Qy 1 MDPLIIFLLIFGTSETQVDYNNVPGTRYDITISSIS-TTYSPTVTLVTNVTKPGPP 59
Db 2 MDHFSFLLLIGTSESQVDYSSPDGTYDITLSSVATTSVSPVSRTLATNVTKPGPP 61
Qy 60 VFLAGERSAGILLSWTTPNPNRIISYIKYKEVCPMOTVYTVQVRKPSDLSLEVLIT 119

Db 62 VFLAGRVSAGILLSWNTPPNPGRGIIISVVKYKEVCPWMTAYTRARAKPDSLEVLIT 121
QY 120 NLNPGTYIIEKVAENSAGIGVSPDPLPOTAESAPGVNLTVEAYNASAVKLIWYLP 179
Db 122 NLNPGTYIIEKVAENNAGIGVSPDPLPOTAESAPGVNLTVEALNYSAVNLWYLP 181
QY 180 QNPKITTSFKISVGHARSIGVVKDVSIRVEDIILTKLPECNENSSBFLWSTASPSPTLGR 239
Db 182 QNPKITTSFKISVGHARSIGVVKDVSIRVEDIILSGKLPECNENSSBFLWSTASPSPTLGR 241
QY 240 VTPPSRTTHSSLTITONEISSVWKEBISFVWTHLAPYTYLFEVSAATTEAGYIDSTIVR 299
Db 242 VTPVTRTQSSSTAARSKISSVWKEBISFVWTHLAPYTYLFEVSAATTEAGYIDSTIVR 301
QY 300 TPESVPEGPQNCVNTGKFSILWDPPTIVTGKFSYRVELYGPSGRILDNSTKOLKF 359
Db 302 TPESVPEGPQNCVNTGKFSISWDPPITIVTGKFSYRVELYGPSGRILDNSTKOLRF 361
QY 360 AFTNLTPFTMYDVIYIAAETSAGTGPKSNI SVTPDPVPGAVFDOLQLAEVESQVRIWKK 419
Db 362 AFTNLTPFTMYDVIYIAAETSAGTGPKSNI SVTPDPVPGAVFDOLQLAEVETEILITWRK 421
QY 420 PRQNGIINQYRKVLVPTGIIILENTLLTGNNEYINDPMAPEIIVNIVPVMVGLYEGSAB 479
Db 422 PRQNGIISQYRKVSVLSTGVVLENTLLTQODESISNPMSPMEIMLVDPMIGFYEGSGE 481
QY 480 MSDDLHSLATFIYNHPDKNFPARBAEDOTSPPVTRNOYITDIAEOLSVYIRBLVPP 539
Db 482 MSDDLHSPASFTYNSHPNDPASTRAEQSSPFWTRNOYITDITABEQSVYVRRLLVPP 541
QY 540 TEHMTSVSAFTMGEPPTVLSVTRTQVPSIIKIINYKNISSSSILLWYDPPEYPNGKI 599
Db 542 TEHMTSVSAFTMGEPPTVLSVTRTQVPSIIKIINYKNISSSSILLWYDPPEYPNGKI 601
QY 600 THYTIYABELDNRAFOITIDNSFLITGLKKYTKYKMRVAASHTHGESLSSEENDIFVR 659
Db 602 THYTIYATELTDNRAFOITIDNSFLITGLKKYTKYKMRVAASHTHGESLSSEENDIFVR 661
QY 660 TSEDEPESPQDVEVDVTADEIRLKWSPPEKPNGIIIAEYVLYKNIIDTYLWKNYSTTDI 719
Db 662 TPEDEPESPQDVQVTVGVSPELRLKWSPEKPNGIIIAEYVLYQNAADTLFVKNYSTTDI 721
QY 720 ILRLNRPHTLXNISVRSYTRFGHGNQVSSILLSVRSSETVDPDAPENITYKNISSGEILS 779
Db 722 IISDLKPYTLXNISRSYTRFGHGNQVSSILLSVRSSETVDPDAPENITYKNISSGEILS 781
QY 780 FLPPSPNGIIEKTYTYLKRSGNEERTINTTSLTONIKVLKXYTOYIIEVSASTLKGBG 839
Db 782 FLPPSPNGIIEKTYTYLKRSGNEERTINTTSLTONIKVLKXYTHYVIEVSASTLKGBG 841
QY 840 VRSAPISILTEEDAPDSPPQDSVKOLSGVTYKLSHQPPELNGIILYTVVWNRSSLK 899
Db 842 IRSRPSILTEEDAPDSPPQDSVKOLSGVTYKLSHQPPELNGIILYTVVWMDKSLR 901
QY 900 TTNVTETSELDLDVNVYSAAVTASTRFGDKTGSNIISFQTPGAPSDPPKOVVYAN 959
Db 902 ALNATEASLVSLDLDVNDYIGACVTASTRFGDNARSSIIINFRTPEGESDPPNDVHYN 961
QY 960 LSSSSILFWTPPPKENGIIQYYSVYRYNTSGTFMQNFTLHETNDPDMNTYSTIIDLKT 1019
Db 962 LSSSSILFWTPPPKENGIIQYYSVYQNTSGTFVQNFLLQVTKESDNVTVSARIYLA 1021
QY 1020 IFSYTYFWLTASTSVGNKSSDIIEVYTDQDIPEGFVGNLYVESISSTAIVNSVPPAQ 1079
Db 1022 IFSYTYFWLTASTSVGNKSSDIIEVYTDQDIPEGFVGNLTAFESISSTAIVHVSPEPQ 1081
QY 1080 PNLGNFYVSLIIQOT-PRHVAPPLYVYERSLYFDNLKXYDILKITEPTEKGSQDVTY 1138
Db 1082 PNLGNFYVSLIIQOT-PRHVAPPLYVYERSLYFDNLKXYDILKITEPTEKGSQDVTY 1141
QY 1139 AOLYIKTEEDVPETSPITINTFKNLSTSVLLSDPPKPNGAIISYDLTLQGENYISFI 1198
Db 1141 AOLYIKTEEDVPETSPITINTFKNLSTSVLLSDPPKPNGAIISYDLTLQGENYISFI 1199

Db 1142 TOLHIKTEEDVPETSPITINTFKNLSTSVLLSDPPKPNGAILGVHLTLQGHANHTFV 1201
QY 1199 TSDNYIILBELSPFTLYSFFAAARTRKGLGPSIILFFYTDESVPPLAPQNLTLINCTSD 1258
Db 1202 TSGNHIVLEELSPPFTLYSFFAAARTRKGLGPSIILFFYTDESAPLAPQNLTLINCTSD 1261
QY 1259 VMLKWSPLPGGIVKVSFKIHEHETDIYKNIISGFKTEAKVGLBPVSTYSIRVSF 1318
Db 1262 VMLTWSPLPGGIVKVSFKIHEHETDIYKNIISGLQTDAKLEGLBPVSTYSIRVSF 1321
QY 1319 TKVGNQFNVVKETTSQESVDDVQNMOCNATSQSVLVKWDPPKXANGIITQWVTV 1378
Db 1322 TKVGNQFNVVKETTSQESVDDVQNMOCNATSQSVLVKWDPPKXANGIITQWVTV 1381
QY 1379 RNSTKVSQDQMYTEIKILLANTSYYFKVRASTASAGEDEBCHVSTLPETVPSVPTNIAF 1438
Db 1382 GNSTKVSQDQMYTEIKILLANTSYYFKVRASTASAGEDEBCHVSTLPETVPSVPTNIAF 1441
QY 1439 SDVQSTSATLWIRPDTILGYFQNYKIITQOLRAQKKEWSECEVEYQKIQYLYEAHLTE 1498
Db 1442 SNVQSTSATLWIRPDTILGYFQNYKIITQOLRAQKKEWSECEVEYQKIQYLYEAHLTE 1501
QY 1499 ETVYGLKKEPRWTRFOVAASTNAGYGNASNIISTKLPDPPDPPENHVHVAATSPSIS 1558
Db 1502 ETVYGLKKEPRWTRFOVAASTNAGYGNASNIISTKLPDPPDPPENHVHVAATSPSIS 1561
QY 1559 NSEPAVITGPTCYLIDVKSVDNDEFNISFISKNEENKTEIEIKDLISFTKYSVVIITAFNG 1618
Db 1562 NSEPAVITGPTCYLIDVKSVDNDEFNISFISKNEENKTEIEIKDLISFTKYSVVIITAFNG 1621
QY 1619 ISAAVEGKSAAEMIVTTLESAPDPPNNMTFQKIPDEVTKFPQLTFLPPSQPNNGIIOVYQ 1678
Db 1622 VSRAVTDGKSSAAEMIVTTLESAPDPPNNMTFQKIPDEVTKFPQLTFLPPSQPNNGIIOVYQ 1681
QY 1679 ALVYREDPTAVQIHNLSIIQKNTFVIAMLEGLKGGTYNISVAVNSAGAPKVPMRI 1738
Db 1682 ALVYREDPTAVQIHNLSIIQKNTFVIAMLEGLKGGTYNISVAVNSAGAPKVPMRI 1741
QY 1739 TMDIKAPARKTKPPIYDATGKLAVTSTTIRMPICVYSDHGPPIKNVQVLAETGAQ 1798
Db 1742 TMDIKAPARKTKPPIYDATGKLAVTSTTIRMPICVYSDHGPPIKNVQVLAETGAQ 1801
QY 1799 HDGNTYKWDAYFNKARPYFTNEGPNPPCTEGKTKFSGNEBIYIIGADNACMIPGNEK 1858
Db 1802 HDGNTYKWDAYFNKARPYFTNEGPNPPCTEGKTKFSGNEBIYIIGADNACMIPGNEK 1861
QY 1859 ICNGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCLISII 1918
Db 1862 ICNGPLKPKQYLFKFRATNMGQTDSDYSDPVKTLGEGLSERTVEIILSVTLCLISII 1921
QY 1919 LIGTAFAPARIRQKKEGGTVSPQDAEIIDTKLDOLITVADLEKDERLTR----- 1972
Db 1922 LIGTAFAPARIRQKKEGGTVSPQDAEIIDTKLDOLITVADLEKDERLTR----- 1981
QY 1973 ---PISKKSFLQHVBEELCTNNLKFQBEFSELKFLQDLSSTDAOLPMNRAKRPNTIKP 2029
Db 1982 SIKPISKKSFLQHVBEELCTNNLKFQBEFSELKFLQDLSSTDAOLPMNRAKRPNTIKP 2041
QY 2030 YNNNRVKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWRWVWETRAKT 2089
Db 2042 Y-NNNRVKLIADASVPGSDYINASYISGLCPNEFIATQGPLGTGDFWRWVWETRAKT 2100
QY 2090 LVMLTQCEKGRIRCHOYWPENDKPVTVFQDITVITKLMEDVOIDMTIRDLKIEREGDCMT 2149
Db 2101 LVMLTQCEKGRIRCHOYWPENDKPVTVFQDITVITKLMEDVOIDMTIRDLKIEREGDCMT 2160
QY 2150 VRQCNFTAWPEHGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2209
Db 2161 VRQCNFTAWPEHGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGTGVFIADHLT 2220
QY 2210 QHINHDVFDYVGLVAELRSEKMCVQNLAQYIFLHQICILLDSNKGSNQPCFVNYSL 2269
Db 2221 QHINHDVFDYVGLVAELRSEKMCVQNLAQYIFLHQICILLDSNKGSGHQPFCFVNYSTL 2280

QY 2270 QKMSLDAMEGDVELEWEETM 2291
DB 2281 QKMSLDAMEGDVELEWEETM 2302

RESULT 12

US-10-314-232-22
; Sequence 22, Application US/10314232
; Publication No. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: UELRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 898
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PTPS31-RD#2
US-10-314-232-22

Query Match 37.8%; Score 4535.5; DB 14; Length 898;
Best Local Similarity 97.1%; Pred. No. 1.7e-253;
Matches 861; Conservative 0; Mismatches 1; Indels 25; Gaps 2;

QY 1429 VPSVPTNIAFSDVQSTSATITWIRPDTILGYFQNYKITTLRAQKCKEWESECEVYQKI 1488
DB 13 VPSVPTNIAFSDVQSTSATITWIRPDTILGYFQNYKITTLRAQKCKEWESECEVYQKI 72
QY 1489 QYLYEAHLTEETVYGLKFRWYFQVAASNAGYGNASNWIKTLLPGPPDGPENHVYV 1548
DB 73 QYLYEAHLTEETVYGLKFRWYFQVAASNAGYGNASNWIKTLLPGPPDGPENHVYV 132
QY 1549 ATSPFSISISWSEPAVITGTCYLIDVKSVDNDEFNIFSKSNEENKTIEIKDLEIFTRY 1608
DB 133 ATSPFSISISWSEPAVITGTCYLIDVKSVDNDEFNIFSKSNEENKTIEIKDLEIFTRY 192
QY 1609 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAPKDPNNNTFOKI PDEVTKFOLTLPSPS 1668
DB 193 SVVITAFGTNISAAYVEGKSSAEMIVTTLESAPKDPNNNTFOKI PDEVTKFOLTLPSPS 252
QY 1669 QPNGNIQVYQALVYREDDPTAVQIHNLISIIQKNTTFVIAMLEGLKGHTYNIISVAVNSA 1728
DB 253 QPNGNIQVYQALVYREDDPTAVQIHNLISIIQKNTTFVIAMLEGLKGHTYNIISVAVNSA 312
QY 1729 GAGKVPKMRITMDIKAPARKPTPTPIYDAGTKLLVTSITITIRMPICYSDDHGPKNV 1788
DB 313 GAGKVPKMRITMDIKAPARKPTPTPIYDAGTKLLVTSITITIRMPICYSDDHGPKNV 372
QY 1789 QVLATETGAQHDGNTVKWDAYFNKARPVTNEGPPNPPCTEGKTPSGNEEIIYIIGADN 1848
DB 373 QVLATETGAQHDGNTVKWDAYFNKARPVTNEGPPNPPCTEGKTPSGNEEIIYIIGADN 432
QY 1849 ACMIPGNEKICNGPLKPKQYLFKFRATNIMQFTDSYSDPKVTLLGEGLSERTVEIIL 1908
DB 433 ACMIPGNEKICNGPLKPKQYLFKFRATNIMQFTDSYSDPKVTLLGEGLSERTVEIIL 492
QY 1909 SVTLCILSIIILGTAIFAFARINQKQEGTYSPODAEIIDTKLKLDQILITVADLEKDE 1968
DB 493 SVTLCILSIIILGTAIFAFARINQKQEGTYSPODAEIIDTKLKLDQILITVADLEKDE 552

QY 1969 RLTR-----PISKKSFLQHVVEELCTNNNLKPFQEFSELPKF 2004
DB 553 RLTRYSSPFFRRKEIFVQILLSYRKSIPKISKKSFLQHVVEELCTNNNLKPFQEFSELPKF 612
QY 2005 LQDLSTADADLPNRAKNEFPNKPNNNNRVKLIADASVPGSDYINASVISGYLCPNEF 2064
DB 613 LQDLSTADADLPNRAKNEFPNKPNNNNRVKLIADASVPGSDYINASVISGYLCPNEF 671
QY 2065 IATQGPLCTVGDVFWRMVWVETRAKTLVMLTQCFEKGIRIRCHOYWPEDNKPVTVFGDIVIT 2124
DB 672 IATQGPLCTVGDVFWRMVWVETRAKTLVMLTQCFEKGIRIRCHOYWPEDNKPVTVFGDIVIT 731
QY 2125 KLMEDVQIDWTIRDLKIERHGDCTVTRQCNFTAMPBHGVPENSAPLIHFVKLVRAHRAHD 2184
DB 732 KLMEDVQIDWTIRDLKIERHGDCTVTRQCNFTAMPBHGVPENSAPLIHFVKLVRAHRAHD 791
QY 2185 TTPMIVHCSAGVGRGTGVTALDHLTOHINDHDPVDIYGLVAELRSMCMVQNLQYIFL 2244
DB 792 TTPMIVHCSAGVGRGTGVTALDHLTOHINDHDPVDIYGLVAELRSMCMVQNLQYIFL 851
QY 2245 HQCILDLNLSKGSNOPICFVNYSALQKMSLDAMEGDVELEWEETM 2291
DB 852 HQCILDLNLSKGSNOPICFVNYSALQKMSLDAMEGDVELEWEETM 898

RESULT 13

US-10-087-684-8
; Sequence 8, Application US/10087684
; Publication No. US20040029116A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: MacDougall, John R.
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David J.
; APPLICANT: Grosse, William M.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catharine E.
; APPLICANT: Casman, Stacie, J.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles E.
; APPLICANT: Gangolli, Esha A.
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-214 CIP
; CURRENT APPLICATION NUMBER: US/10/087,684
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: CuraSeqList version 0.1

```
; SEQ ID NO 8
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-684-8

Query Match      37.4%; Score 4488; DB 15; Length 855;
Best Local Similarity 99.6%; Pred. No. 9e-251;
Matches 850; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1053 PEGFVGNLTYESISSTAINVSWPPAQNGLVFFYVSLILQOTPRHVRPPLVYERSIYF 1112
Db      |||||
QY      |||||
Db      |||||
QY 1113 DNLEKTYDILKTPSTEGFSDTYTAQLYIKTEEDVPETSPINTFNKLSSTVLLSWD 1172
Db      |||||
QY      |||||
Db      |||||
QY 1173 PPVKPENGAIISYDLTLQGPENYSFITSNDYIIIEELSPFTLYSPFAAARTKGLGPSSI 1232
Db      |||||
QY      |||||
Db      |||||
QY 1233 LFFYTDSEVPLAPPQNLTINCTDFVWLKWSPLPGGIVKVSFKIHEHETDTIYKN 1292
Db      |||||
QY      |||||
Db      |||||
QY 1293 ISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMCMATS 1352
Db      |||||
QY      |||||
Db      |||||
QY 1353 WSVLVKWDPPKANGIIQYVMTVERNSTKVSPODHMTFFIKLLANTSIVFKVRASTSA 1412
Db      |||||
QY      |||||
Db      |||||
QY 1413 GGEDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQ 1472
Db      |||||
QY      |||||
Db      |||||
QY 1473 KCKEWESECEVYQKIYLYEAHLTEETVYGLKFRWYFQVAASTNAGYGNASNIWSTK 1532
Db      |||||
QY      |||||
Db      |||||
QY 1533 TLPGPDGPPENHVVATSPFSSISWSBPAVITGPTCYLIDVKSVDNDEFNISFTKSNE 1592
Db      |||||
QY      |||||
Db      |||||
QY 1593 ENKTIKLEIFTRYSVVITAFPTGNIISAAAYVEGKSSAEMI VTTLESAPKPPNNMTFOK 1652
Db      |||||
QY      |||||
Db      |||||
QY 1653 IPDEVTKFQITLPPSQPNQNTQVQALVYREDDPTAVQIHNLIIQKNTFTVIAMLEGL 1712
Db      |||||
QY      |||||
Db      |||||
QY 1713 KGSHTYNI SVYAVNSAGAKVPMRITMDIKAPAREKTKPTPIYDATGKLLVTSITITIR 1772
Db      |||||
QY      |||||
Db      |||||
QY 1773 MPICYYSDDHGPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPYFTNBFNPPCTEKG 1832
Db      |||||
QY      |||||
Db      |||||
QY 1833 TFPSGNEEYIIGADNACMIPGNEDKICNGPLKPKQYLFKFRATNIMQFTDSYSDPV 1892
Db      |||||
QY      |||||
Db      |||||
QY 1893 KTLGEGLSERTVE·1905
Db      |||||
QY      |||||
Db      |||||
QY 1905 KTLGEGLSERTLE 855
Db      |||||
QY      |||||
Db      |||||

; FILE REFERENCE: 21402-214
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; CURRENT APPLICATION NUMBER: US/10/218,779
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,-926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-218-779-8

Query Match      37.4%; Score 4488; DB 15; Length 855;
Best Local Similarity 99.6%; Pred. No. 9e-251;
Matches 850; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1053 PEGFVGNLTYESISSTAINVSWPPAQNGLVFFYVSLILQOTPRHVRPPLVYERSIYF 1112
Db      |||||
QY      |||||
Db      |||||
QY 1113 DNLEKTYDILKTPSTEGFSDTYTAQLYIKTEEDVPETSPINTFNKLSSTVLLSWD 1172
Db      |||||
QY      |||||
Db      |||||
QY 1173 PPVKPENGAIISYDLTLQGPENYSFITSNDYIIIEELSPFTLYSPFAAARTKGLGPSSI 1232
Db      |||||
QY      |||||
Db      |||||
QY 1233 LFFYTDSEVPLAPPQNLTINCTDFVWLKWSPLPGGIVKVSFKIHEHETDTIYKN 1292
Db      |||||
QY      |||||
Db      |||||
QY 1293 ISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVQNMCMATS 1352
Db      |||||
QY      |||||
Db      |||||
QY 1353 WSVLVKWDPPKANGIIQYVMTVERNSTKVSPODHMTFFIKLLANTSIVFKVRASTSA 1412
Db      |||||
QY      |||||
Db      |||||
QY 1413 GGEDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTLQRAQ 1472
Db      |||||
QY      |||||
Db      |||||
QY 1473 KCKEWESECEVYQKIYLYEAHLTEETVYGLKFRWYFQVAASTNAGYGNASNIWSTK 1532
Db      |||||
QY      |||||
Db      |||||
QY 1533 TLPGPDGPPENHVVATSPFSSISWSBPAVITGPTCYLIDVKSVDNDEFNISFTKSNE 1592
Db      |||||
QY      |||||
Db      |||||
QY 1593 ENKTIKLEIFTRYSVVITAFPTGNIISAAAYVEGKSSAEMI VTTLESAPKPPNNMTFOK 1652
Db      |||||
QY      |||||
Db      |||||
QY 1653 IPDEVTKFQITLPPSQPNQNTQVQALVYREDDPTAVQIHNLIIQKNTFTVIAMLEGL 1712
Db      |||||
QY      |||||
Db      |||||
QY 1713 KGSHTYNI SVYAVNSAGAKVPMRITMDIKAPAREKTKPTPIYDATGKLLVTSITITIR 1772
Db      |||||
QY      |||||
Db      |||||
QY 1773 MPICYYSDDHGPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPYFTNBFNPPCTEKG 1832
Db      |||||
QY      |||||
Db      |||||
QY 1833 TFPSGNEEYIIGADNACMIPGNEDKICNGPLKPKQYLFKFRATNIMQFTDSYSDPV 1892
Db      |||||
QY      |||||
Db      |||||
QY 1893 KTLGEGLSERTVE·1905
Db      |||||
QY      |||||
Db      |||||
QY 1905 KTLGEGLSERTLE 855
Db      |||||
QY      |||||
Db      |||||

; SEQUENCE 8, Application US/10218779
; Publication No. US20040029222A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William
; APPLICANT: Alsobrook II, John
; APPLICANT: Lepley, Denise
; APPLICANT: Rieger, Daniel
; APPLICANT: Burgess, Catherine
; APPLICANT: Caeman, Stacie
; APPLICANT: Spytek, Kimberly
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy, Suresh
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar
; APPLICANT: Vernet, Corine
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Malyankar, Uriel
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles
; APPLICANT: Gangolli, Esha
; FILE REFERENCE: 21402-214
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; CURRENT APPLICATION NUMBER: US/10/218,779
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,-926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-218-779-8

RESULT 14
US-10-218-779-8
```

Db 243 ISGFKTEAKUGLEPEVSTTSIRSAFTKVGNGQFNSNVKFTTQESVDPDVQNMOCWATS 302
Qy 1353 WQSVLVKWDPPKANGIIITQYMTVVERNSKVSQDQDMYTFIKLLANTSIVFKVRASTSA 1412
Db 303 WQSVLVKWDPPKANGIIITQYMTVVERNSKVSQDQDMYTFIKLLANTSIVFKVRASTSA 362
Qy 1413 GEGDESTCHVSTLPETVPSVPTNIAPSDVOSTSATLTWRPDTILGYFQNYKITTLQRAQ 1472
Db 363 GEGDESTCHVSTLPETVPSVPTNIAPSDVOSTSATLTWRPDTILGYFQNYKITTLQRAQ 422
Qy 1473 KCKEWESEECVEYQKQYLYEAHLTETVYGLKFRWYRFOVAASNAGYGNASNIWK 1532
Db 423 KCKEWESEECVEYQKQYLYEAHLTETVYGLKFRWYRFOVAASNAGYGNASNIWK 482
Qy 1533 TLPGPPDPGPPENHVHVVATSPFISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKSNE 1592
Db 483 TLPGPPDPGPPENHVHVVATSPFISISWSEPAVITGPTCYLIDVKSVDNDEFNIFSIKSNE 542
Qy 1593 ENKTIKOLEIFTRYSVVIPTAGNISAAVYEGKSAEMIVTTLSSAPKDPNNMTFQK 1652
Db 543 ENKTIKOLEIFTRYSVVIPTAGNISAAVYEGKSAEMIVTTLSSAPKDPNNMTFQK 602
Qy 1653 IPDEVTKFOLTFLPBPSPNGNIQYQALVYREDPTAVQIHNLISIIOKNTVFVIAMLEGL 1712
Db 603 IPDEVTKFOLTFLPBPSPNGNIQYQALVYREDPTAVQIHNLISIIOKNTVFVIAMLEGL 662
Qy 1713 KGGHTYNISSVYVNSAGAGKVPWIRITMDIKAPARKTKPTPTPYDATGKLLVTSITITIR 1772
Db 663 KGGHTYNISSVYVNSAGAGKVPWIRITMDIKAPARKTKPTPTPYDATGKLLVTSITITIR 722
Qy 1773 MPICYSDHGHPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGK 1832
Db 723 MPICYSDHGHPIKNVQVLATETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGK 782
Qy 1833 TKESGNEEYIIIGADNACMIPGNEKICNGPLKPKQYLFKFRATNIMGQFTDSDYSDPV 1892
Db 783 TKESGNEEYIIIGADNACMIPGNEKICNGPLKPKQYLFKFRATNIMGQFTDSDYSDPV 842
Qy 1893 KTLGEGLSERTVE 1905
Db 843 KTLGEGLSERTLE 855

RESULT 15
US-10-314-232-15
; Sequence 15, Application US/10314232
; Publication No. US20030138932A1
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/10/314,232
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/361,096
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-314-232-15

Query Match 15.9%; Score 1902.5; DB 14; Length 401;
Best Local Similarity 92.8%; Pred. No. 1.1e-101;
Matches 361; Conservative 2; Mismatches 1; Indels 25; Gaps 2;

Qy 1927 FARIROKQKEGGTYSPQDAEIIIDTKLQDLITVADLELKDRLTR----- 1972
Db 14 YSEIROKQKEGGTYSPQDAEIIIDTKLQDLITVADLELKDRLTRYSSFFFRKEIFVI 73
Qy 1973 -----PISKKSFLQHVVELCTNNLKFQEESELPKFLQDLSSDADLPWNRKN 2022
Db 74 QLLSVKSKIKPISKKSFLQHVVELCTNNLKFQEESELPKFLQDLSSDADLPWNRKN 133
Qy 2023 RFPNIKPYNNNRVVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTVDGFWRMV 2082
Db 134 RFPNIKPYNNNRVVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLGTVDGFWRMV 192
Qy 2083 WETRAKTLVMTQCFEKGIRCHQWYWPEDNKPVTVPDGIIVITKLMEDVQIDMTIRDLKIE 2142
Db 193 WETRAKTLVMTQCFEKGIRCHQWYWPEDNKPVTVPDGIIVITKLMEDVQIDMTIRDLKIE 252
Qy 2143 RHGDCMTVROCNFTAWPEHGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGVF 2202
Db 253 RHGDCMTVROCNFTAWPEHGVSPENSAPLIHFVKLVRAASRAHDTTPMIVHCSAGVGRGVF 312
Qy 2203 IALDHLTQHINDHDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLILSNKGSNQPIC 2262
Db 313 IALDHLTQHINDHDFVDIYGLVAELSERMCMVQNLAQYIFLHQICILDLILSNKGSNQPIC 372
Qy 2263 FVNYSAIQKQDSDLAMEGDVELEWETTMM 2291
Db 373 FVNYSAIQKQDSDLAMEGDVELEWETTMM 401

Search completed: June 30, 2005, 15:20:57
Job time : 251 secs

THIS PAGE BLANK (USPTO)